DTIC FILE COPY

(2)

NAVAL POSTGRADUATE SCHOOL
Monterey, California

AD-A209 473

DTIC
ELECTE
JUN 29 1989
S E D

THESIS

CONVERSION OF HARD-COPY DOCUMENTS
TO DIGITAL FORMAT UTILIZING
OPTICAL SCANNERS AND OPTICAL STORAGE MEDIA

by

Robert Ryland Taylor

March 1989

Thesis Advisor: Barry A. Frew

Approved for public release; distribution is unlimited

89 6 26 026

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY 2b DECLASSIFICATION / DOWNGRADING SCHEDULE	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution is Unlimited
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) Code 37	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b. ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
8a. NAME OF FUNDING / SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)

CONVERSION OF HARD-COPY DOCUMENTS TO DIGITAL FORMAT UTILIZING OPTICAL SCANNERS AND OPTICAL STORAGE MEDIA

12 PERSONAL AUTHOR(S)

Taylor, Robert Ryland

13a. TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM ____ TO ____	1989 March	120

16. SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Optical Scanners, OCR, ICR, Image Scanners, Optical Disks, CD-ROM, WORM, Optical Information System

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

Storage of hard-copy archival paper documents requires vast amount of storage space and time to search and retrieve. Technology exists today to convert hard-copy texts using optical scanners and storing in a digital formal on optical disks. This thesis conducts an indepth current technology research of optical scanners, optical storage mediums, and optical information systems. Utilizing the thesis documents presently stored in the library aboard Naval Postgraduate School, as a statistical base, this research analyzes the requirements to convert the thesis documents to digital format. This thesis concludes that an image optical information system is a viable alternate to storing hard-copy documents and recommends follow-on thesis research to build an in-house optical information system.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Barry A. Frew	22b TELEPHONE (Include Area Code) (408) 646-2924 22c OFFICE SYMBOL Code 54FW

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

U.S. Government Printing Office: 1986—606-243

UNCLASSIFIED

Approved for public release; distribution is unlimited.

Conversion of Hard-Copy Documents to Digital Format Utilizing Optical Scanners and Optical Storage Media

by

Robert R. Taylor
Lieutenant Commander, United States Navy
B.S., University of Utah, 1977

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

NAVAL POSTGRADUATE SCHOOL
MARCH 1989

Author: Robert R. Taylor

Robert R. Taylor

Approved by: Barry A. Frew

Barry A. Frew, Thesis Advisor

Carl R. Jones, Second Reader

David R. Whipple, Chairman
Department of Administrative Sciences

K. T. Marshall

Kneale T. Marshall
Dean of Information and Policy Sciences

ABSTRACT

Storage of hard-copy archival paper documents requires vast amount of storage space and time to search and retrieve. Technology exists today to convert hard-copy texts using optical scanners and storing in a digital formal on optical disks.

This thesis conducts an indepth current technology research of optical scanners, optical storage mediums, and optical information systems.

Utilizing the thesis documents presently stored in the library aboard Naval Postgraduate School, as a statistical base, this research analyzes the requirements to convert the thesis documents to digital format.

This thesis concludes that an image optical information system is a viable alternate to storing hard-copy documents and recommends follow-on thesis research to build an in-house optical information system.

INSP

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

TABLE OF CONTENTS

I. INTRODUCTION

A. DISCUSSION

The reference room at the Knox Library aboard Naval Postgraduate School is reaching its maximum storage capacity in terms of shelving its archive documents. Technology exists today to convert hard-copy documents using optical scanners and storing in a digital format on optical disks.

> Companies are finding economic as well as technical virtues to optical-disk technology that justify going optical. Some firms can cost-justify these systems by the space they save versus all other data storage media. The value of a document may be beyond measure, but a square foot of a floor space occupied by a filing cabinet certainly has its price. (Alter, 1988, p.18)

Converting to digital format will require less storage space and provide for a faster search and retrieval capability. An optical-disk based information system made by LaserData of Lowell, Mass. was installed at the Maine Medical Center in Portland, Maine which allowed the hospital to clear an entire floor - 7,200 square feet - of a building dedicated to medical records and radiology records. The hospital was out of storage space, so they were looking to recapture space, rather than go out and build new space.

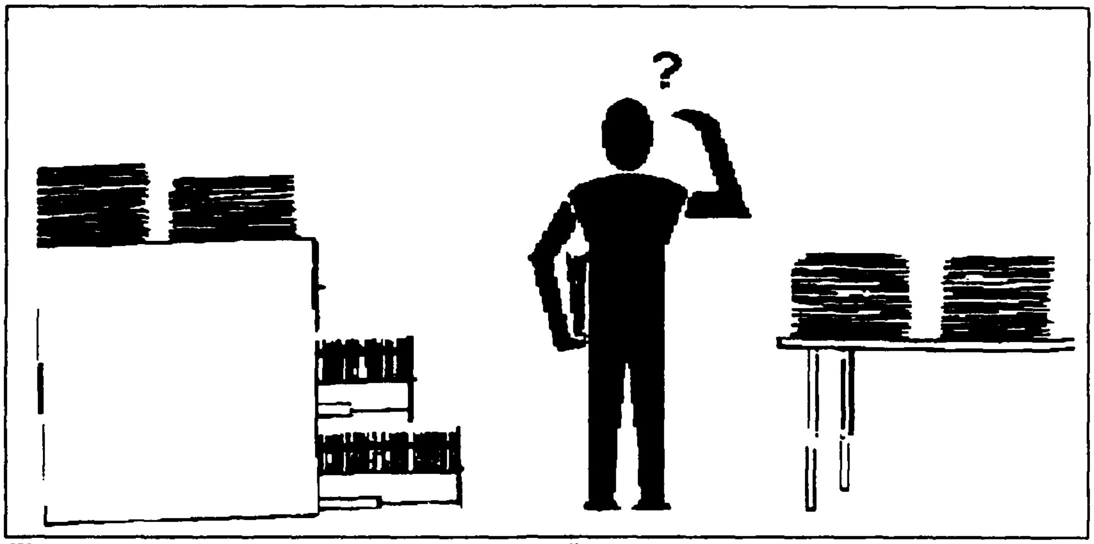

Figure 1 Today's paper filing system. ("Wang Laboratories, Imaging - Primer Series," 1987, p. 4)

In Portland, building new storage space would have cost $100 per square foot. Recouping 7,200 square feet at $100, equals a savings of $720,000. That was the primary cost-justification. Another was being able to instantly find the records, which improved overall patient care. (Alter, 1988, p. 18)

Another example of current utilization, USAA (United Services Automobile Association) set up a 1,300 workstation document-processing system to be shared by over 2,000 employees in its property and casualty policy-service operation.

According to Charles A. Plesums, manager of image systems, the company began processing 2 percent of that

> operation's document workload in mid-July 1988 and will expand to 100 percent by early 1989.
>
> Seven years from now, he added, optical disks will store 300 million pages and save 39,000 square feet of space. (Lesher, 1988, p. 33)

This thesis will consider the analysis and design of an Optical scanner to Optical storage medium Document Processing System.

B. SCOPE

This thesis includes an indepth review of current Optical Scanner and Optical Storage Medium technologies presently available. The purpose of this review to provide the reader the different options available for designing an Optical Information System.

Analysis of requirements for implementing an archive Optical-disk-based document processing system will be conducted. Alternative systems and solutions will be addressed and a recommendation will be submitted for possible implementation.

C. METHODOLOGY

Utilizing the thesis texts presently stored in the library as a statistical population, a small sample will be

used to conduct research on Optical Scanners. Questions to be answered concerning Optical Scanners will be 1) What is the time required to convert text to data, 2) How accurate is converting hard copy text to digital format, and 3) What are the digital storage requirements. Review of Optical Storage Media that will best match the requirements of an Optical Scanner will be addressed. Presently there is thesis research being conducted in the area of Indexing an Optical Disk using Hypertext, and Storage requirements using Optical Disk. This thesis will primarily emphasize Optical Scanners and the initial phase of conversion in an Optical information system.

II. OPTICAL SCANNERS

A. INTRODUCTION

The initial phase of converting hard-copy text to digitized format in an Optical Information System is scanning the document either by an Optical Character Reader or by an Image scanner.

Today modern scanners can combine the functions of reading text and processing image information because they contain more, and more complex components and algorithms than did earlier scanners. No scanner yet exists that can scan a page and interpret text and graphics in a single pass. However, software now exists that provides the option of utilizing an image scanner to scan for either text or graphics in a single pass and then combining the two to produce a digitized copy of the original.

B. ELEMENTARY CONCEPT OF SCANNERS

A scanner obtains optical information (about light and dark areas on the image) from the original image. Next, the electronic converter units translate that optical image information into digital information. A processor unit

manipulates the digital data according to specified instructions, in order to create an output image that can be, in some way, different from the original image. Figure 2 illustrates the conversion from a scanned image to digital format for comparison. The reader unit of a scanner combines a light source, several mirrors, and a lens. These components illuminate the original image and reflect light from it. More light is reflected from lighter areas of the original than from darker areas.

A photoconverter converts information about the reflected light into an electrical voltage. An analog-to-digital converter further changes the electrical (analog information into a digital (binary) data format.

The digital data is passed to an image processor, where it can be manipulated to produce the desired output. In the image processor, adjustments may be made to the size and shape, resolution, and contrast of the output image.

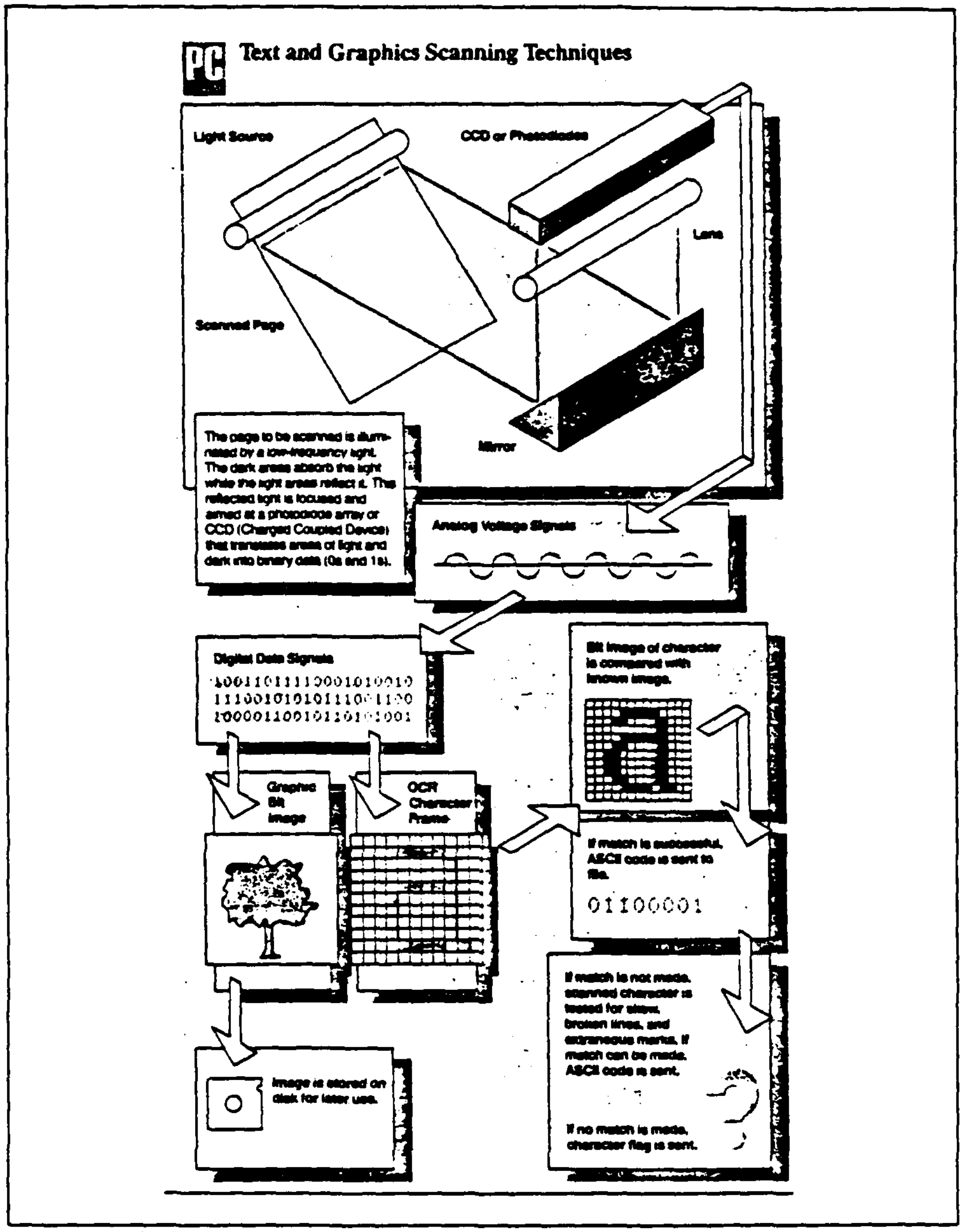

Figure 2 Text and Graphics Scanning Techniques (PC Magazine, 1986, p. 134)

C. INDEPTH REVIEW OF HOW SCANNERS WORK

1. The Light Path

The original image is first illuminated by a light source in the reader unit of a modern scanner. Light reflected from the original image is passed by mirrors to a lens, which focuses the reflected light and passes it from the reader unit to the converter units.

Figure 3 illustrates the path that the light takes from the original image to the CCD (Charged Coupled Device), or photoconverter.

The above listed steps of the scanning process are essentially the same for both text and image processing. However, depending on the scanner's design, either the page is moved over a fixed scanning element or the scanning element is moved over a fixed page. Most OCR's have a fixed scanning element and most image scanners have a moving scanning element.

The light source of the scanning element may be a laser, or another type of high intensity lamp. For an image scanner, the light source is mounted on a carriage, so that it moves to illuminate the original image, not all at once,

but in a systematic manner. The carriage moves in the slow scan direction, illuminating a strip of the original image with each movement. While a strip is thus illuminated, the fast scan occurs.

a. Slow Scan

In the slow scan direction the light source moves stepwise to each strip of the original image, where it pauses while the fast scan takes place. The distance it moves is dependent on the resolution setting. This distance corresponds to the height of a pixel in the output image.

b. Fast Scan

In the fast scan direction the light source pauses for a brief interval. Information from the illuminated strip of the original image is read and converted into digital data before it is processed. The illuminated strip is divided into discrete sections. The width of each section is determined by the resolution. This width corresponds to the width of a pixel in the output image.

Light reflected from the original image is thus divided into discrete areas which are processed separately. Each is represented as a pixel in the output image.

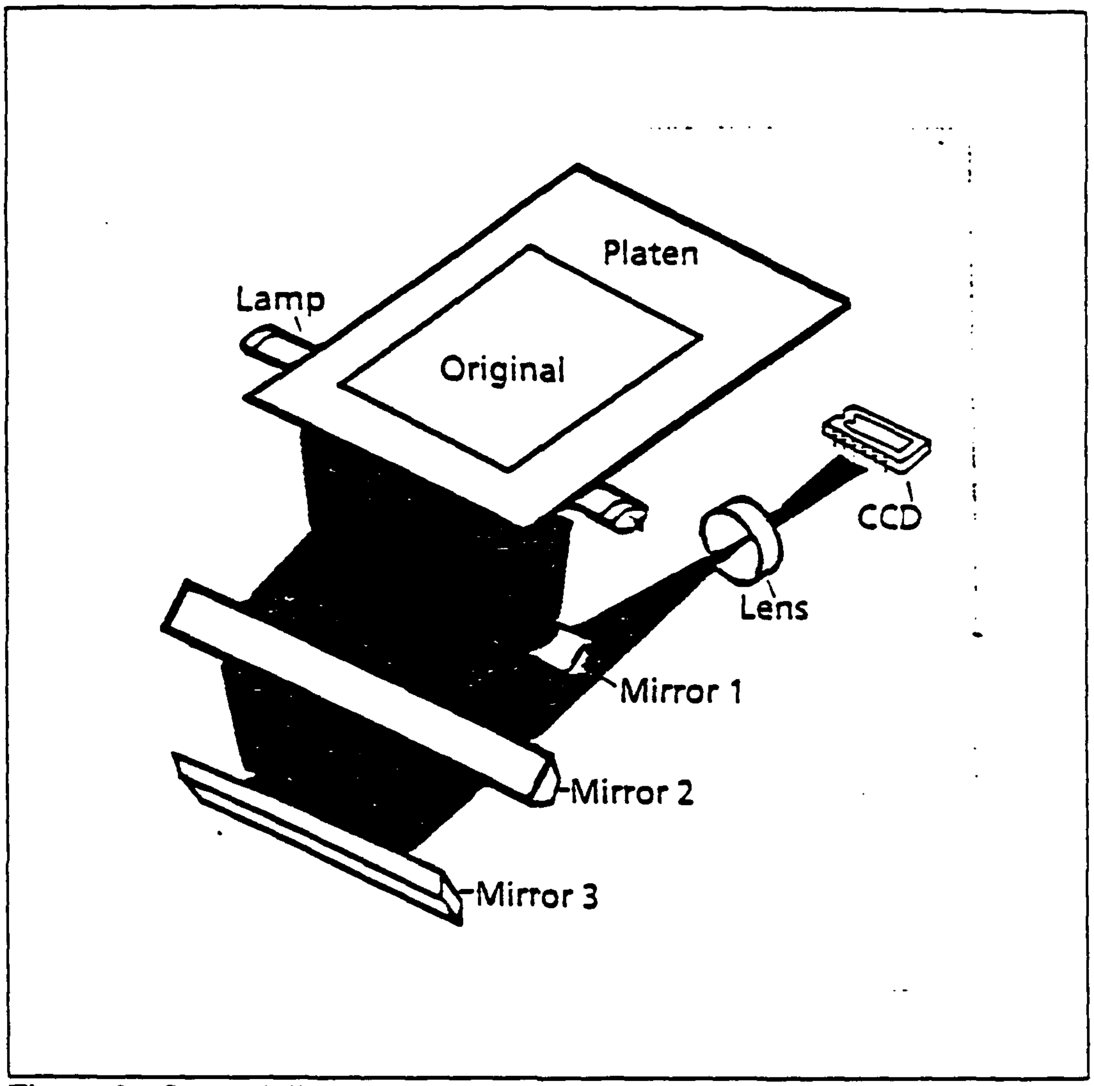

Figure 3 Scanned light path to the CCD. (XEROX 7650 reference manual, 1987)

A series of mirrors pass the reflected light from the original image to the lens. In this way the focal length of the reflected image is effectively made longer.

The longer focal length permits the use of a relatively small lens.

The lens, like the lens in a camera, focuses the reflected light from the final mirror onto a specific site (corresponding to a pixel) on the surface of the photoconverter.

2. Light Signal to Digital Signal

The converter units of a scanner contain electronic devices which convert, or transform, the information reflected from the original image into electronic data.

As the light reflected from the original image is passed to the photocells on the surface of the CCD, they convert that optical signal into an electrical signal (a Voltage) proportional to the "size" of the optical signal. The "size" of the optical signal is the amount of reflected light. That is, a white area of the original image reflects more light, so it generates a greater voltage.

The electrical signal requires one more transformation before its information can be understood by the processor. An analog-to-digital converter performs this

final transformation, and passes the digital signal to the processor.

From this point on, the data manipulation processes for optical character readers and image scanners are very different.

D. OPTICAL CHARACTER READERS

1. What is OCR?

An OCR is defined either as optical character readers or as optical character recognition used in the process of converting an image of text into computer readable form (i.e., ASCII).

The original concept of an OCR was a device that could only digitize characters produced by a typewriter. A new acronym, ICR, is being used by some vendors to replace OCR. ICR, defined either as Internal Character Recognition or Intelligent Character Recognition, includes the capability to recognize omni-font characters or otherwise known as the many different fonts and characters produced by todays computers and printers. For the purpose of this thesis, OCR will be used to imply both OCR and ICR.

OCR is accomplished by analyzing the image of a character and then deciding what character the image represents. Unfortunately, OCR is not an exact science and consequently, any OCR process is inherently imperfect. Recognition errors will occur, regardless of the particular OCR technology.

2. OCR Technology

Once the scanned image is converted into digital data, OCR scanners digitize the characters a line at a time and then isolate them, character by character, into frames ranging from 24 by 40 pixels up to 30 by 50 pixels. The individual frames are stored in RAM for character recognition processing.

There are two broad categories of character recognition processing commonly used in today's OCR/ICR scanners. The first, and perhaps the oldest, is commonly called Matrix Matching or Template Matching. The second, a more recent development, is referred to as Feature Extraction.

a. *Matrix Matching*

Matrix Matching, in its simplest form, can be thought of as comparing the image of an unknown character with images of known characters and finding the nearest match. Figure 4 illustrates how a scanned image is compared to a template.

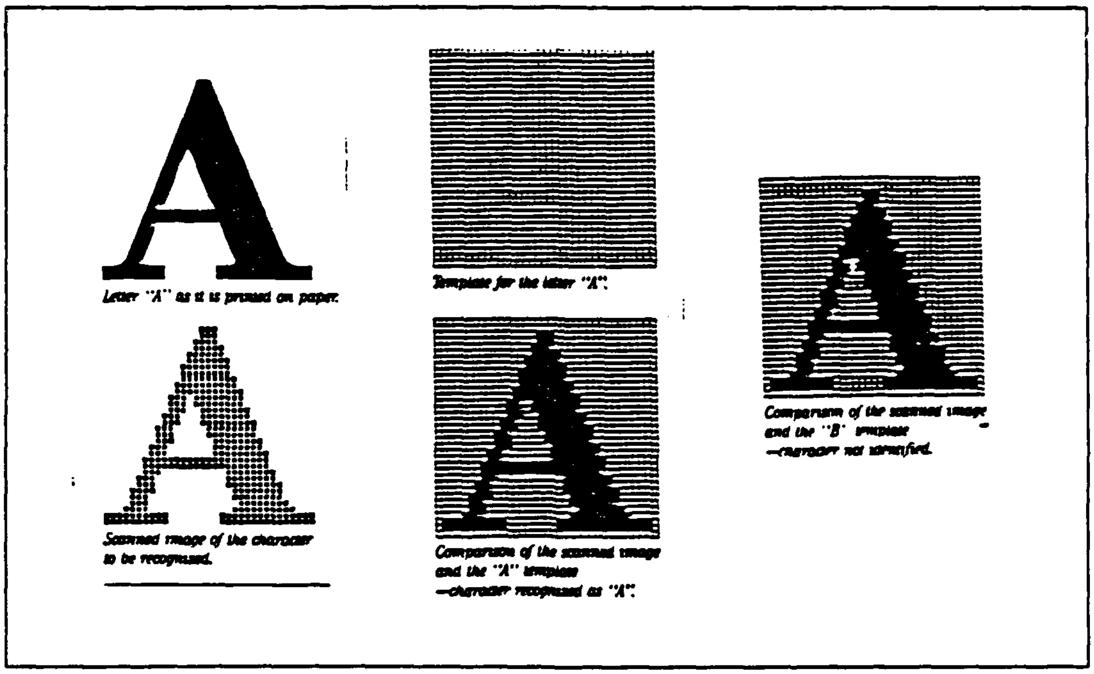

Figure 4 Template Matching. (MICRO User's Guide, 1988, p. 20)

The very nature of this technique requires a complete set of templates for each font the system will read. This means that multi-font matrix matching systems need considerable memory for the font libraries.

Another disadvantage associated with matrix matching is its sensitivity to minor variations in fonts. Two fonts that look the same to a user may not be recognized equally well in matrix matching due to subtle differences in character shapes or sizes. On the positive side, matrix matching is relatively insensitive to broken characters, which occur all too frequently in ordinary documents.

b. Feature Extraction

The term Feature Extraction is used in the industry to describe any OCR technique other than Matrix Matching. As a result, the name does not convey much information about how OCR is being done. Of the feature extraction techniques in use today, the most popular is Topological Feature Analysis.

Topological Feature Analysis involves identifying the important features of a character image and, based on these features, deciding what character the image represents. These features can include vertical strokes, horizontal strokes, line endings, closed curves, open curves, slanted strokes, intersection of strokes, et cetera. Figure 5 illustrates the comparison between a scanned image and the primitive features extracted.

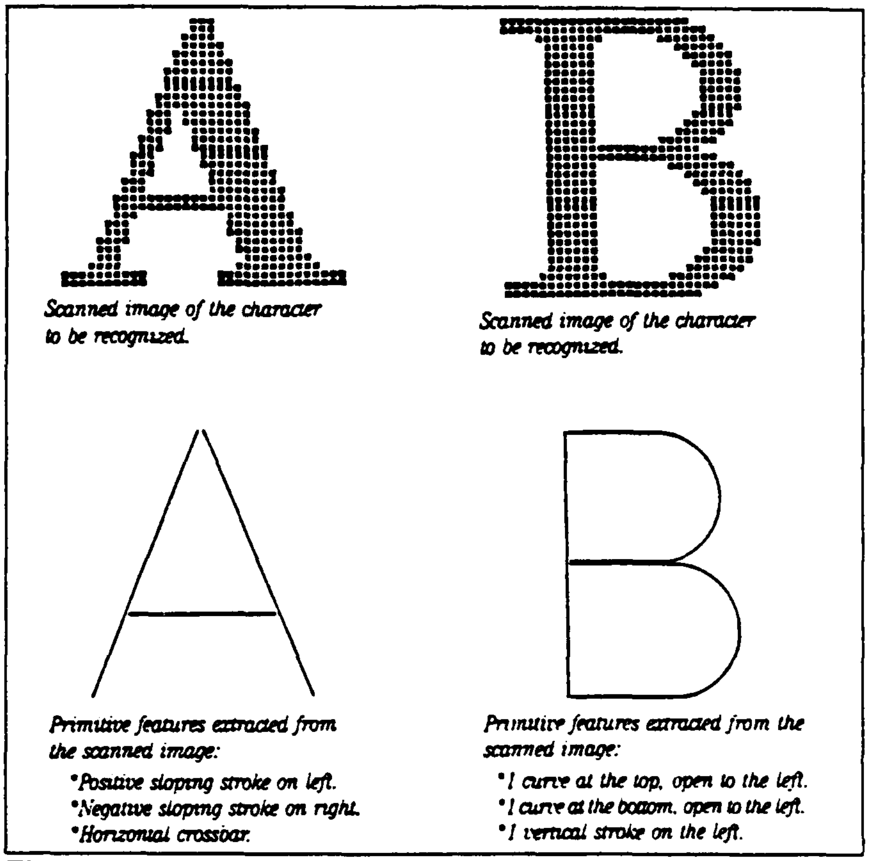

Figure 5 Feature Matching. (MICRO User's Guide, 1988, p. 23)

The nature of this technique makes it relatively insensitive to slight variations in character shape and size. Another advantage to feature extraction techniques is that, in most cases, less memory is required for the font libraries. For example, the features of the letter "e" are fairly constant for a wide variety of typefaces. A

disadvantage of feature extraction is its sensitivity to broken characters.

E. IMAGE SCANNERS

Image scanners work like laser printers in reverse. A scanner converts image information into electrical signals that can be stored in a computer, whereas a laser printer converts the image into charges on the surface of a photosensitive drum.

The electrical signals from the scanner's CCD (which reads an entire row at a time) are converted to numeric values and stored in RAM until the entire image is scanned. Figure 6 illustrates the digital data captured from a single strip of the letter "T" (one pixel in height) as it is read by the scanner.

This data is stored in RAM like a two-dimensional mosaic of dots that represents the original image. This two-dimensional mosaic is otherwise known as a bit map.

Two parameters are very important in image scanning: resolution, usually expressed in pixels per inch (PPI), and the number of levels of grayscale information captured for each pixel.

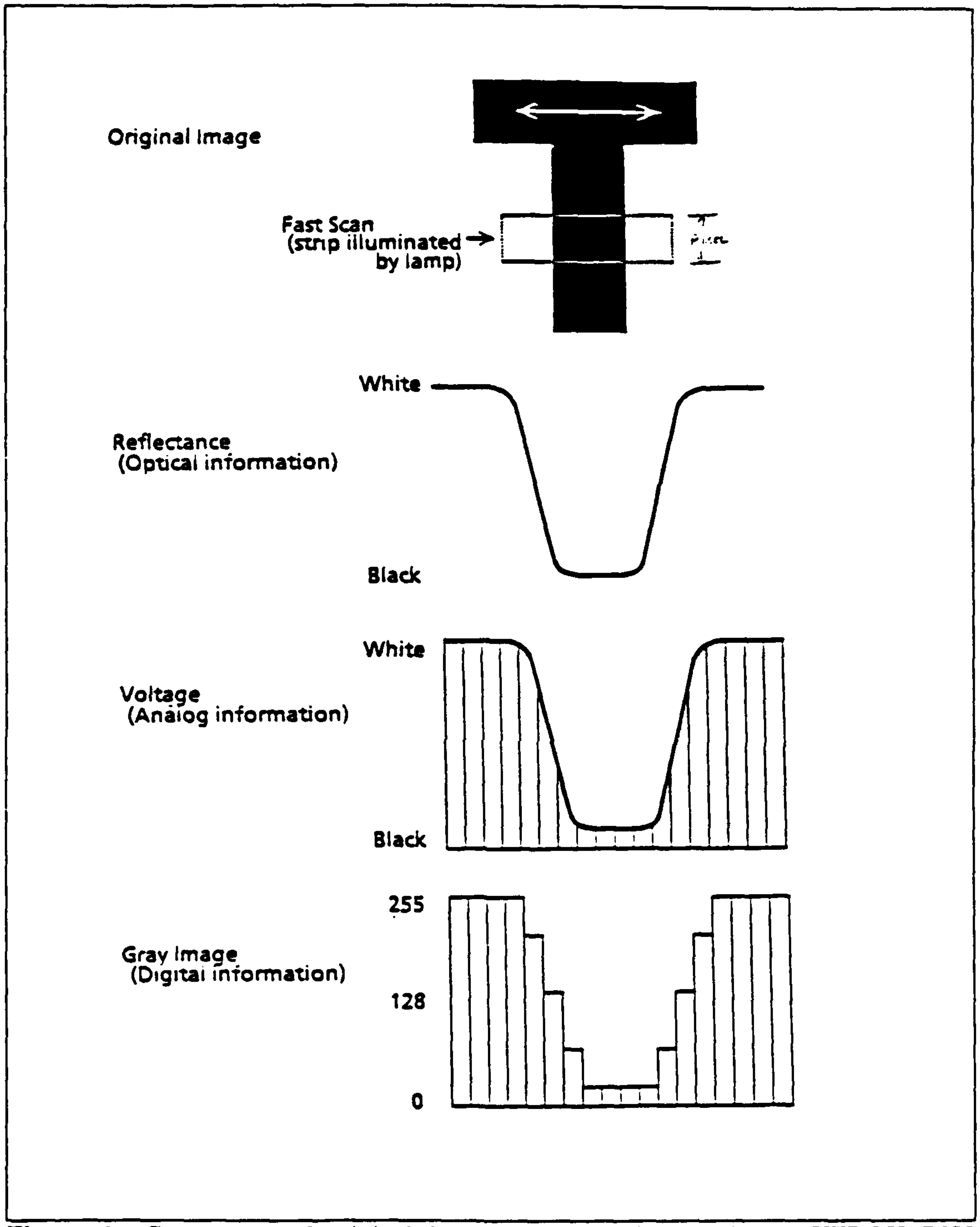

Figure 6 Conversion of original image to digital information. (XEROX 7650 Reference Manual, 1987)

1. Resolution

Resolution is defined as the number of pixels read or displayed per inch (PPI), both horizontally and vertically.

Most image scanners have resolutions up to 300 PPI, meaning that a single scanned line across the width of an 8- by 10-inch image contains 2,400 pixels. And if the lengthwise resolution is also 300 PPI, then there will be 3000 pixels in a single column the length of the image. In all, it takes about 7.2 million pixels to represent an 8- by 10-inch image. An image that size requires approximately 1 megabyte of storage.

Increasing the resolution allows more detail (finer lines or sharper changes in gray in an image) to be resolved and improve the appearance of a scanned image. Figure 7 illustrates the difference in appearance of the letter "a" at different resolutions.

2. Levels of Grayscale

If you think of a typical fine grain photograph, the number of shades of gray that can be reproduced are essentially infinite, at least as far as the human eye can

see. However, in the digital world there is a limit to the number of shades of gray. Everything must be represented in discrete steps and it takes more information (bits of data) to represent more steps. Thus 4 bits of data are required to represent 16 levels of gray per pixel, 6 bits to represent 64 levels, and 8 bits to represent 256 levels.

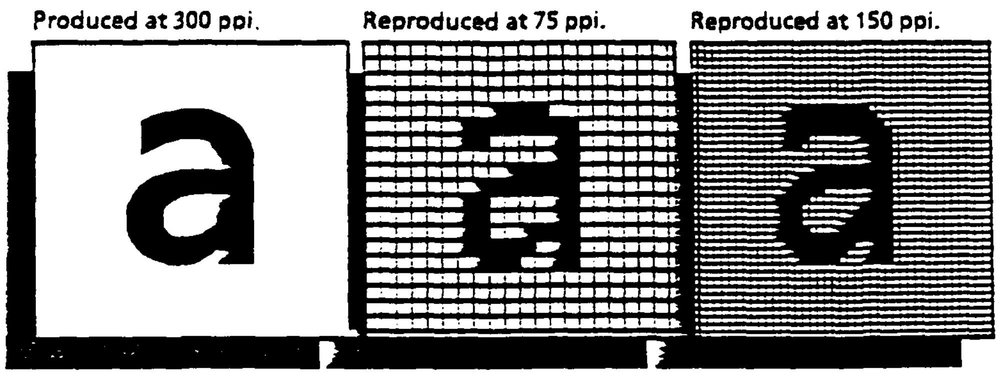

Figure 7 Comparison of different PPI settings. (XEROX 7650 Reference manual, 1987)

The higher the resolution of an image and the higher the number of levels of grayscale the image contains, the higher the quality of the image. Unfortunately, the size of image files increases with respect to the square of the resolution and linearly with respect to the number of bits of grayscale information. One scanned image page could easily require several Mbytes of storage as illustrated by

Table I. Additionally, processing and retrieval time increases as the size of the stored data increases.

F. COMPRESSION/DECOMPRESSION OF A SCANNED IMAGE

Because of the extremely large memory requirements of scanned images, it was necessary to develop compression/decompression techniques to increase the number of pages that could be stored on a particular device.

Table I FILE SIZE (IN MEGABYTES) FOR AN 8 1/2 X 11 INCH SCANNED IMAGE. (DESKTOP PUBLISHING, FALL 1988, P. 29)

Resolution of scanned image (ppi)	Grayscale Levels			
	2	16	64	256
300	1.1	4.2	6.3	8.4
400	1.9	7.5	11.2	15.0
600	4.2	16.8	25.3	33.6

Until compression/decompression chips became available, image compression was performed in software, taking at least 30 seconds for a typical page. Now with image compression/decompression processor chips, such operations take only a few seconds.

Most image compression processors are based on the CCITT group 3 and 4 standards, developed for use in facsimile transmission. These standards are based on a combination of two image compression algorithms, known as Modified Huffman (MH) and Modified Read (MR) encoding. (Matlin, 1988, p. 75)

1. Modified Huffman Encoding

Also known as one-dimensional encoding, MH works on an image one horizontal line at a time. Each run length, or continuous string, of black or white pixels is given a code base on the probability of that particular length. To achieve image compression, the codes for the most probable run lengths must be shorter than the run lengths themselves.

The CCITT Group 3 standard employs MH encoding. In this algorithm, the codes representing the document run lengths are selected from one of two 64-element tables representing black and white run lengths of 0 to 63 pixels. These tables were derived from statistics based on eight standard documents, which are available from the CCITT. For longer run lengths, make-up code tables exist for run lengths in multiples of 64 pixels.

As an example, if an entire run length of 8 1/2 inches is white, and the document is scanned at 300 PPI, a white run of 2544 pixels is indicated (rounded off to the next-lowest byte). This run length can be represented by a make-up code for 2496 pixels (000000011110), a white run-length code for 48 pixels (00001011) and an end-of-line code (000000000001). Since only 32 bits are required to represent the original 2544 pixels, a compression ratio of 79.5:1 is achieved for this line. Of course, this is a particularly easy line to compress.

2. Modified Read Encoding

Also known as two-dimensional encoding, MR coding takes advantage of the vertical correlation between adjacent lines within a document. It has been estimated that 50 percent of all transitions from white to black, or vice versa, occur directly below a transition on the previous line. To encode using the MR algorithm, the relationship between a transition on the current line and the previous line is determined. If the current line transition is within three pixels of a transition on the previous line, a vertical mode is indicated. This case is represented by a

short code indicating vertical mode, and another code indicating the relative distance between the current line transition and the transition above it. If the distance between transitions is more than three pixels, the pixel distance is encoded using the appropriate MH code. This is known as horizontal mode. A third technique, known as pass mode, is used to realign the transition pointers between the coding and reference lines.

III. OPTICAL STORAGE MEDIA

A. INTRODUCTION

Before optical storage, it was difficult to have video, audio, image, and text data on-line because of the large memory required to store the various types of data. With the advent of optical storage media, different forms of information can be digitized, integrated, and displayed as a single form of information.

The extremely high-density recording capability of optical devices enables one 5 1/4 inch optical disk drive to store 654 million bytes (654 megabytes) of information. That's equivalent to the amount of data contained on 1800 360K floppy disks or 33 (20 megabyte) hard disks or 260,000 pages of text. A single 12-inch optical platter can store as much as four GB's (gigabytes) of information. Four GB's or four billion bytes is equal to the data stored in 160 file cabinets or the amount of data stored on 120 2,400-foot magnetic tapes (Dukeman, 1988, p.82). Larger discs are available containing even larger amounts of data, for example Eastman Kodak Company recently introduced an optical system that can store more than a terabyte of information.

One terabyte is equal to a trillion bytes. The Kodak system 6800 uses 14 inch optical discs that store 6.8 gigabytes (billion bytes) each of randomly accessible information. The automated library unit can accommodate as many as 150 discs (and 150 times 6.8 billion yields a figure in excess of one trillion). ("CD-ROMS: The Laser's Edge in Data Storage", 1987, p. 53)

Compared with magnetic disks and tapes, optical media is almost indestructible. Optical disks can be mailed without special precautions, and taken through X-ray machines and airport scanning devices. Optically stored data is unaffected by the environment or magnetic fields. Some optical media last for 30 to 100 years, but magnetic media has an average life expectancy of only three to five years.

Optical disks are removable and thus the data can be securely stored. Optical disks don't stretch over time as do magnetic tapes. Most optical media can't be altered, and optical media is less expensive per megabyte of storage. Data access time for optical disks is still slower than magnetic disks, but as the product matures and proliferates in its target markets of data distribution, publishing,

database archiving, and imaging data, access time will improve. Paper-intensive environments should trade increased access times for large capacity, unattended backup capability and high-volume storage of integrated data, text and images. (Levine, R., 1988, p. 50)

B. ELEMENTARY CONCEPT OF OPTICAL DISKS

Information is recorded on a plastic-coated glass disc in the form of pits and lands. Pits are indented 0.12 micrometer deep and 0.6 micrometer wide into the surface of the disc. Flats measure between 0.9 and 3.3 micrometers in length.

Data on the CD-ROM disk is arranged in a spiral pattern, radiating from the center toward the outer edge. A space of 1.6 micrometers separates the lines of data in the spiral. This configuration yields an effective track density of 16,000 tracks per inch. In contrast, floppy disks have a density of 96 tracks per inch.

Before data is inscribed on the disc, it must first be translated into a special dialect of the binary channel code that is used to transfer data between more familiar magnetic formats and computer devices. In magnetic tape formats, the

ones and zeros represent digital information. CD-ROM channel code assigns ones to mark movement from a land to a pit or a pit to a land, zeros mark a continuation of lands or pits in series.

A low power laser is used to read data from the disc surface. Light rays are aimed by an optical head over the information track on the spinning disc, and the amount of light reflected back to the optical head indicates the presence of a flat, which reflects more light, or a pit, which reflects less light. The series of flats and pits of digital data unscrambles the code into data the computer can use. ("CD-ROMs: The Laser's Edge in Data Storage", 1987, p. 52)

C. CD-ROM: COMPACT DISK-READ ONLY MEMORY

CD-ROM offers prerecorded optical storage. It's a read-only device; you can read the information on the disk, but it can't be altered. Used to distribute a common database that doesn't have to be updated constantly to multiple division, departments or branch offices, it ensures that the data is protected against tampering or accidental erasing and is ideal for archival purposes.

Most systems of this type store 600 MB's (megabytes) of data on a 4.7-inch CD-ROM disk and drive. Half-height 5 1/4-inch drives also are available. CD-ROM manufacturers have embraced the High Sierra Group and ISO 9600 standard for file organization, thus the 4.7-inch drive has become the industry standard. (Levine, R., 1987, p. 50)

CD-ROM is the most economical optical media for mass distribution of databases. The cost of preparing a master disk is relatively expensive, but after mass production, the cost per copy can be as low as $2 in a large-scale distribution.

For CD-ROM, optical disks are mass produced regardless of whether the encoded data represent video, audio or text. Once the information has been transcribed into digital format and the special cue codes have been added, all the data is transferred to a master tape.

Once the information is recorded on the plastic-coated glass disc, the glass disc is used to create a metalized master disc. The surface of the master is transferred onto nickel shells to form negatives and positives from which 'stamper' copies are made for mass replication. The stamper

is used to transfer the information onto nickel shells with reflective aluminum and then covered with lacquer.

D. WORM: WRITE-ONCE, READ-MANY

These optical storage devices permit one-time writing but unlimited reading of data and images. Although you can't overwrite or erase previously stored data, you can update it by writing new information into a file at another location on the disk. The new file then is linked to the original file through software and is retrieved in its place. This operation is transparent to the user.

WORM optical technology consists of a high-intensity laser beam that heats and permanently changes the surface of the disk as it writes and stores information on the disk. The writing process, which varies from vendor to vendor, ultimately results in a change in reflectivity of the information layer of the disk. Figure 8 illustrates how a WORM drive works.

Most WORM drives use a glass- or plastic-based substrate to enclose a sensitive recording layer. Eastman Kodak Co. uses an aluminum substrate in its 14-inch Optical Disk System 6800.

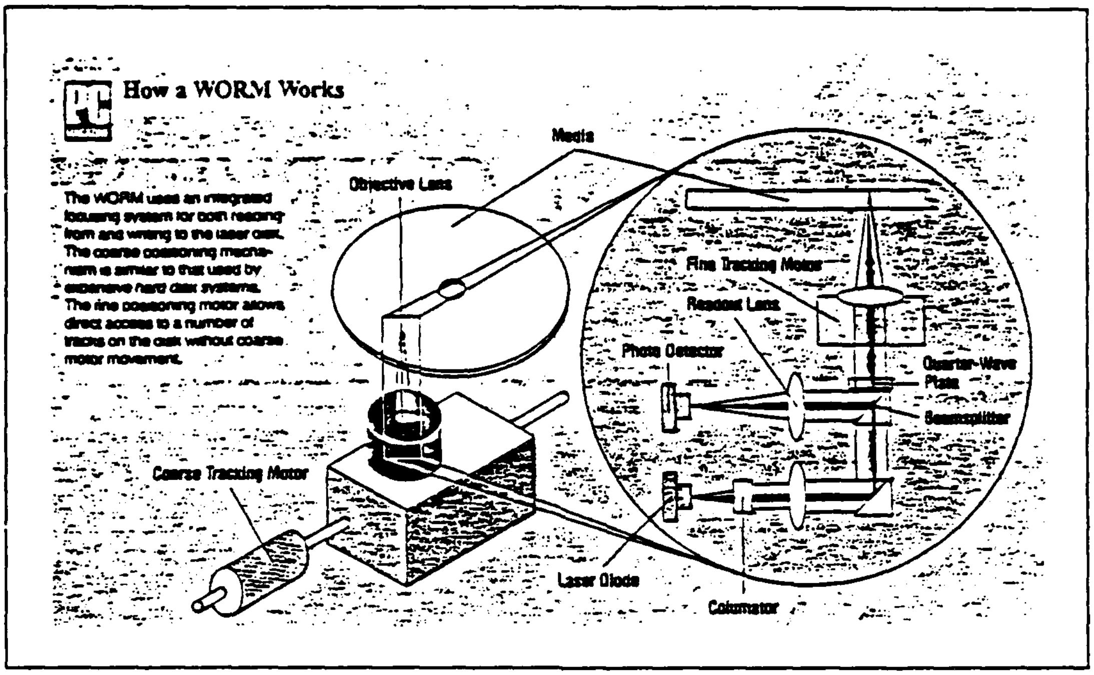

Figure 8 How a WORM drive works. (PC Magazine, June 1987)

Presently, three recording technologies are used for WORM optical disks; ablative, vesicular and phase-change.

1. Ablative Recording

Ablative technology stands out as the most common method of writing data. This technology, also known as pit-forming, burns a hole in the active layer of the media.

2. Vesicular Recording

Vesicular technology, also known as bubble-forming, heats the media until it melts and forms a bubble, or explosion, of the polymers on the active layer of the media.

3. Phase-Change Recording

Phase-change technology, actually produces a change in the media from a crystalline to an amorphous state.

E. MAGNETO-OPTICAL DISKS: ERASABLE OPTICAL STORAGE

Magneto-Optical disks provide the same capability of storing and retrieving data as present magnetic drives do, but with the storage capability 12 to 50 times the amount of data currently packed on magnetic hard disk drives.

Magneto-Optical disks drives use a combination of technologies to store and retrieve information. They rely upon materials whose particles can be magnetically oriented either up or down but whose orientation can't be changed easily at normal temperatures.

Storing information on the disk is performed by a strong laser beam, as illustrated in Figure 9, which heats a microscopic spot in a multi-layered material sealed in the rotating disk. When the temperature of the magneto-optical layer reaches a certain point, its magnetic orientation can be changed easily by a magnetic field in the drive.

After the laser beam is removed, the exposed disk region retains its magnetized orientation.

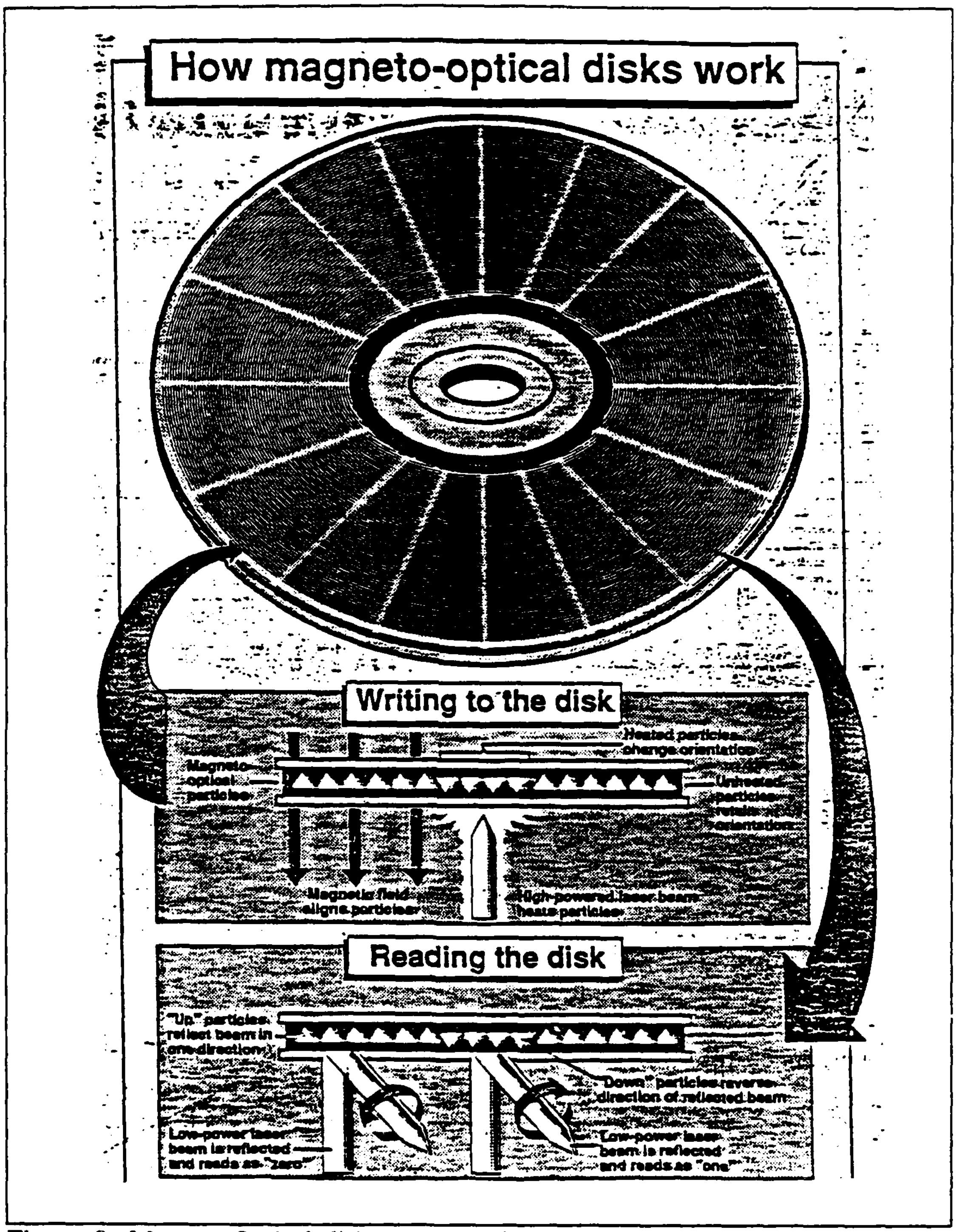

Figure 9 Magneto-Optical disk concept. (Infoworld, 1988)

To record information, a first pass is made with the laser and magnetic field to erase an entire section of the recording surface by orienting all the spots the same way, to represent zeros. Then, a second pass is made with the magnetic field reversed, but this time the laser heats only spots to be changed from zero to one.

To read the information later, as illustrated in Figure 8, a weaker, polarized laser beam is shone at the spot. Depending on the direction of the magnetization of the recording layer, the polarization of the beam rotates 180 degrees, a phenomenon known as the Kerr effect.

After striking the surface, the polarized beam is reflected back to a photodetector, which reads the variations. With the stronger beam, the information can later be "erased" by again heating the spot and altering the magnetic orientation.

F. SUMMARY

When designing an archival information system, the optical media of choice is WORM. Cost of producing the master disk would be prohibitive for CD-ROM unless there was a distribution base to make it cost effective. Magneto-

optical is still being development, but does provide an alternative to WORM if there is a need to erase the original disk. However, magneto-optical disks are more expensive than WORM disks, leaving WORM as the economic medium of choice for archival purposes.

IV. OPTICAL INFORMATION PROCESSING

Optical information processing is a relatively new technology, utilizing optical storage media to store the data created by a document data processing system.

A. DOCUMENT DATA PROCESSING

What is Document Data Processing? Document Data Processing is the procedure of converting information stored on paper to digitized format. Document data processing includes the ability to electronically store, retrieve and reproduce the original information contained on paper.

Document data processing systems in the past have used optical character readers to convert paper information to electronic format. Microfiche or magnetic storage devices were used to store the electronically converted data.

In the past, the high expense and relative low capacity of magnetic media have precluded its use for storing archival quantities of documents in other than character coded format. (Kapoor, 1988, p. 28)

Before the discovery of optical disk, it was impractical to maintain images on-line because of the large memory

requirements of storing a single page. (Grigsby, 1988, p. 62)

With the advancements in optical storage media and the techniques to compress images into manageable sizes, technology exists today to design a document data processing system that will merge and manage diverse forms of information, including image, text, alphanumeric data and voice.

B. OPTICAL INFORMATION SYSTEM

Optical information processing systems provide both an image and a data processing solution. These digital systems utilizing optical storage media to store, and retrieve are the missing link in the integration of paper documents, microfilm, computer data, and word processing text. This technology provides solutions not previously available to solve information access and distribution requirements associated with a total information transaction. (Grigsby, 1988, p. 60)

Optical information systems are an idealistic alternative to document data processing systems. Optical

disks not only store images and data, but also the retrieval software for index data management.

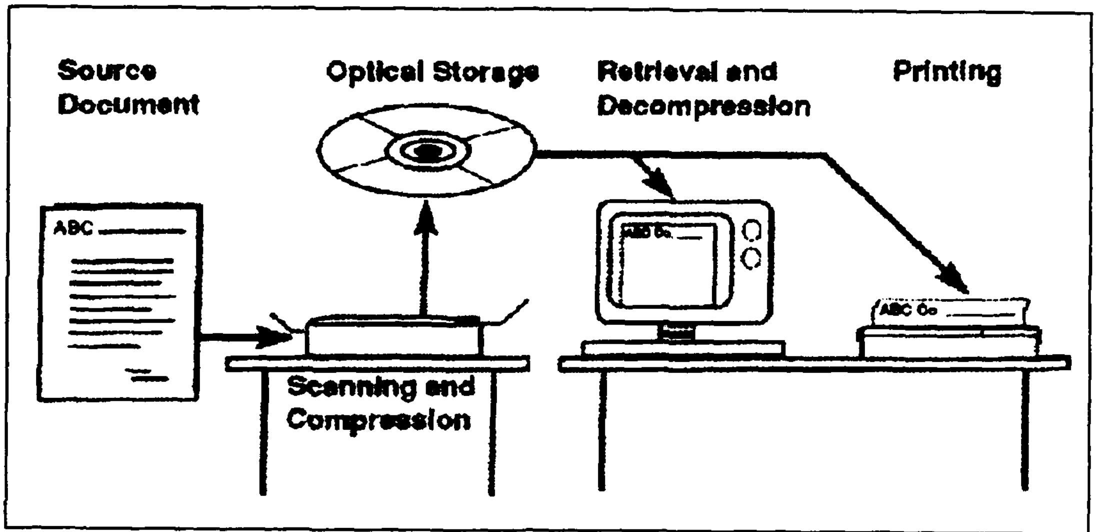

Figure 10 Optical Information System.

A basic stand alone PC-based optical information processing system consists of an image scanner to create digitized images of documents, workstations utilizing a bit-mapped high resolution monitor to display the images, magnetic hard disks to store indexed information and act as a buffer before storing the digitized image, optical disks to store and retrieve the images and a laser printer to reproduce the image.

These stand alone systems are modular and can be expanded into large, organization wide, document image and

data management systems with multiple workstations and optical disk storage libraries.

When the optical storage media is connected to a network several people can review the stored document simultaneously, eliminating the delays that result from passing a paper file from person to person. This process allows the actual paper file to be stored in a low cost, secure area. Paper files, which are subject to theft and accidental loss, often must be sent out for review. Each journey risks the integrity of the file and costs additional time and expense.

An optical storage system may include jukeboxes. Optical jukeboxes use robotics to mount and dismount a large number of optical disks. A jukebox may contain as many as 95 disks and up to five separate drives, yielding quick access to over three million images. When an image is requested, the correct optical disk is robotically selected and mounted and the desired image displayed in seconds.

C. BENEFITS OF OPTICAL INFORMATION PROCESSING

A primary reason for using computer paper, computer output microfiche, and magnetic tape archival storage is low

cost. Until now, it has been too expensive to keep massive amounts of data "on-line". Optical information systems provide on-line desktop delivery of current, as well as historical, computer information. (Grigsby, 1988, p. 62)

1. Huge Capacity/Space Savings

Dozens of comparisons have been made to dramatize the optical disk systems ability to compress volumes of paper onto a disk. This ability provides high-density on-line storage at a comparatively low cost as demonstrated by the following table:

Table II STORAGE CAPACITY AND COST COMPARISON OF DIFFERENT STORAGE MEDIA. (MICROSOFT PRESS, 1986)

MEDIA	Floppy Disk	Hard Disk	Magnetic Tape	CD-ROM	WORM	Large Optical
Capacity (in MB)	.36-1.2	5-50	30-300	540-680	200-300	1,000 -4,000
Cost per MB	1093.59	63.63	54.64	2.48	17.40	21.41

2. Speed of Retrieval

"In a current environment, even in a hurry, it may take 10 minutes to retrieve a paper document, copy it, add notes, and fax it to a remote office," says Michael Florio,

vice president Document Technology, INC. (Dukeman, 1988, p. 82). With an optical information processing system, the process would take only 30 seconds, drastically reducing the clerical functions of research, filing and hard-copy reproduction. Another example: with an optical disk jukebox storing 280 gigabytes, a document can be accessed in 7 to 10 seconds (Dukeman, 1988, p.82).

3. Shared Access/Remote Availability

Information on paper can be in only one place at a time, or it can be copied and multiplied beyond control. In an optical information system only one copy of an image exists, but users have access on an "as needed" basis.

4. File Integrity

File integrity is another significant reason that automation of paper documents is important. Many documents are simply lost or not available due to misfiling or out-of-file situations. Using WORM optical disk the document cannot be misplaced or altered once it has been written. Additionally, when connected to a network, several people can review the stored document simultaneously, eliminating

the delays that result from passing a paper file from person to person.

5. Archival Life

Storage life of optical media is estimated to be more than 30 years, it can be duplicated on new media at more frequent intervals. The life expectancy of optical disks far surpasses the estimated life of 5 to 10 years for magnetic storage. At the 1988 AIIM show in Chicago, Sony Corporation announced that accelerated tests showed that Sony WORM media is capable of a one-hundred-year life (Dukeman, 1988, p. 84).

6. Cross Reference Indexing

Once identified by a multiple level cross reference index, images can be retrieved by a number of desired fields. Indexing multiple key fields allows greater flexibility for accessing data.

7. No Head Crashes

Unlike magnetic disks, optical disks do not experience head crashes.

8. Distribution

Coupling optical systems with a communication device, users can send and receive documents in seconds. By adding a facsimile capability to a system, this enables the system to send a document image to virtually anywhere there is another fax machine.

V. METHODOLOGY AND DATA

A. INTRODUCTION

In search of hardware to evaluate for an Optical information system, the researcher used three sources to acquire data from: 1) In-place operational systems, such as the EDS Deers Enrollment Processing Center which uses a 3M Docutron 9000 optical information system and the Defense Language Institute Graphics Department which uses a Kurzweil 4000 optical character recognition system, both in Monterey Ca.; 2) Companies that specialize in Optical information system integration, such as TAB Products Co. Palo Alto, Ca., Anamet Laboratories, Inc. Hayward, Ca., LaserData Inc., Lowell, Ma., and Wang Laboratories, Inc., Lowell, Ma.; and 3) Vendors that sell either OCR scanners or Image scanners, such as Xerox or Western Office Supply in Santa Clara.

Prior to answering the research questions, a summarization of findings of both optical character readers and image scanners will be discussed.

B. OPTICAL CHARACTER READER EVALUATION

With the purpose of fulfilling the need of converting thesis documents into digital format, several OCRs, including the latest technology advanced OCRs available on the market, were used for evaluation.

Two top of the line OCR scanners were evaluated. The Kurzweil 5000 was demonstrated by Western Office Supplies of Santa Clara, Ca. and the Calera CDP 3000 was demonstrated by Anament Laboratories, Inc. of Hayward, Ca. Both have self contained processors and are designated as Omni-font readers. Both have Automatic Document Feeders (ADF) capable of processing 50 pages at a time. With their built-in processors, both were able to background scan, while permitting the PC to perform other functions.

Additionally, True Scan, also a product of Calera, was demonstrated by Western Office Supplies. True Scan uses an image scanner, an extended RAM board added to a PC and software to perform OCR/ICR. True Scan does not have the capability to perform background scanning.

Optical character readers were originally designed to read text only. Current models advertised the ability to

read a page and distinguish between graphics and text. In a sense this was true, distinguishing several fonts of text and ignoring any form of graphics. Selection to scan either in an image mode or text mode had to be made prior to scanning.

If a single page contained both text and graphics, the page would need to be scanned once for text and once for graphics and then to obtain a digitized copy of the original page, the text and the bit map image of graphics would need to be merged.

This process may work well in an office environment where only a few documents a day might be digitized. But when converting large document databases, the time to preview each page prior to scanning, scan the page twice if needed, and merge the text with graphics would be too time intensive to be practical for a large conversion project.

Optical character recognition is still not an exact science. It is the opinion of the researcher that the recognition capability of todays models is vastly improved over earlier models, but there were still numerous errors made by all models previewed.

Newer models can now distinguish text printed in columns and around graphics but still have great difficulty with formats. A full page of text from a thesis averaged 2 to 3 character recognition errors. Recognition errors were easy to correct, but it was the experience of the researcher, that the time required to correct format errors was more intensive.

Appendix A contains pages from the original thesis used for research. Appendix B contains the unedited results using the same pages in Appendix A and scanned with a Kurzweil 5000. The Kurzweil 5000 did an excellent job of reading text, such as the small print on page 1 of the thesis document DD Form 1473. But it illustrates how time intensive it would be to correct the recognition errors and to reformat the page in an acceptable form for permanent storage.

C. IMAGE SCANNER EVALUATION

Several image scanner models were reviewed. The only noticeable difference between the various models as illustrated by Table III was the amount of time it took to

scan a standard 8 1/2 x 11 page. The resolution and grayscale quality was comparable among all models.

The Fujitsu Model M3094E was used by TAB Products and Anamet Laboratories in the integration of their optical information systems. The Fujitsu model was rated fastest among several commercially available models, averaging 7 seconds per page at 200 PPI resolution.

Table III TIME COMPARISON REQUIRED TO SCAN A SINGLE PAGE.

SCANNER	TIME TO SCAN A SINGLE PAGE (resolution = 200 PPI)
Hybrid (3M Docutron 9000 system)	< 1 sec
Fujitsu M3094E	7 sec
Microtek MS 300A	24 sec
XEROX 7650	31 sec

D. RESEARCH QUESTIONS

Utilizing an original copy of a thesis, and the optical character recognition and image scanners discussed above, the following questions were addressed:

1. What is the time required to convert text to data?

a. Optical Character Recognition scanners

Time required to convert text to data depended upon the processor of the individual scanner selected. The scanners with the greater built-in processing capability, such as the Kurzweil 5000 or the Calera CDP 3000 averaged 30 seconds or less per page. Using the True Scan board attached to an Image scanner, the average time was 60 seconds per page. The Kurzweil 4000, using 1985 technology averaged 90 seconds per page.

Times mentioned do not include the time required to correct the errors, nor the time it would take to rescan the page as graphics and attempt to combine the two. Both error correction and combining graphics could take up to an additional 5 minutes per page dependent upon the number of errors and format of the graphics.

b. Image Scanners

Time required to convert an image to digital format was dependent upon the individual processor tested and the resolution selected.

Times for the different processors ranged from less than 1 second per page (scanning both sides) for the

hybrid scanner designed for the 3M Docutron 9000 system to 31 sec per page for the XEROX 7650 image scanner. Comparisons were based on scanning at 200 PPI. (See Table III)

Decreasing or increasing the resolution had the same effect on time to scan. For the XEROX 7650 decreasing the resolution to 75 PPI decreased the time to scan to 14 seconds. Increasing to 300 PPI required 38 seconds and increasing to 400 PPI required 154 seconds. The same page was used for the above time analysis, which demonstrates the increased time required when increasing scanning resolution.

2. How accurate is converting to digital format?

If the document was strictly text the accuracy rate was quite high for OCRs scanners. They rarely had more than 2 or 3 errors a page, but if there was any form of graphics such as figures or tables, the error rate went up drastically.

For Image scanners, the accuracy is a matter of resolution. For a resolution of 75 or 100 PPI, the quality was good but generally not as good as the original, with some of the smaller details harder to read. 200 PPI

resolution was just as good or better than the original. 300 or 400 PPI resolution produced a product that was much better than the original. Appendix C demonstrates the quality difference between pages scanned at 200 and 300 PPI.

3. What are the digital storage requirements?

Text scanned by optical recognition readers required very little storage space as compared to image scanned text. The entire 8 pages read by the Kurzweil 5000 in Appendix B, required only 15,171 bytes to store.

Image scanning requires vastly larger amounts of memory. An uncompressed page scanned at 300 PPI requires an approximate 1 megabyte of memory. A compressed page at 300 PPI still requires approximately 40,000 bytes of storage space. Table IV provides a sample of the compressed file sizes required for individual pages scanned in Appendix A.

Table IV FILE SIZES REQUIRED FOR INDIVIDUALLY SCANNED PAGES.

Page #	200 PPI	300 PPI
Cover	13,717	20,588
1	50,405	76,600
4	24,207	36,361
17	15,526	23,834

VI. ANALYSIS

A. THESIS DOCUMENT STORAGE REQUIREMENTS

The area in the Naval Postgraduate School's Knox library where the thesis documents are stored, is referred to as the thesis cage. It is so named because of the wire walls that enclose the area to control access.

The thesis cage comprises an area of 23 feet by 21 feet and a ceiling height of 8 feet. This small area, utilizing compact shelving, contains approximately 23,500 theses. For each thesis title stored, there is one hard bound and one soft bound document. Therefore, there are approximately 11,750 original thesis documents dating back to the early 50's. Each quarter there is an additional 200 to 250 new theses produced. Adding approximately 1000 new thesis documents annually.

Selecting 20 theses documents at random, the average size of a thesis was 108.3 pages in length, of which 27.9 pages were graphs or charts and 7.4 pages were pictures. It is important to note that each thesis contained approximately 25 percent graphics of some form. For this reason OCR was not considered a viable alternate for

converting the thesis documents and therefore will not be considered in the continuation of the analysis.

For a four month period from September to December 1988, the thesis cage was used, on the average, 4.2 times per day. (The above average did not include Sundays and school breaks between quarters when the library was not being utilized to its fullest capacity). With this in mind, one optical disk processing work station would be more than adequate to fulfill the needs for search and retrieval.

Available on the market today are 1.6 gigabyte per side 12 inch optical WORM disks, with two gigabytes per side and greater being evaluated for market introduction.

Using 12 inch optical disks and a recommended 200 PPI scanning resolution to replace the 11,750 thesis documents, it would take approximately 33 gigabytes or 10.3 optical disks to store all theses currently in the cage. 1000 new thesis documents would require 2.8 gigabytes or approximately one new 12 inch disk each year.

B. COST ANALYSIS

1. Optical information system cost analysis.

A system that fulfills the requirement for an optical information system for the library is the TAB Laser-Optic Filing System 2000. A single desk measuring eight feet in length contains the complete system. The system integrates the following components; one image scanner, one high resolution monitor, one cpu and hard disk, one 12 inch optical disk drive, and a laser printer.

The cost of the TAB Laser-Optic Filing System 2000 with the 12 optical disk drive is $69,950. Each 12 inch optical disk costs $575. Initial purchase would require 11 optical disks to store the entire thesis library, plus two additional disks to cover the first two years of expected additional thesis documents. Technology is expected to increase storage capacity of the 12 inch optical disk, so more than two years in advance purchase is not recommended. The cost of purchasing the 13 optical disks is $7475. Therefore, the initial hardware/software cost to implement the system is $77,125.

2. Hard-copy to digital conversion cost analysis.

Conversion cost is determined by time and cost for an individual to complete the conversion.

Conversion time consists of 1) the time to prepare the document for scanning, i.e., such as removing staples, 2) the time to scan the document, 3) the time to index the document for storage on the optical disk, and 4) the time required to actually store the document on the disk.

The image scanner of the TAB Laser-Optic Filing System 2000 can scan a page in an average 7 sec at 200 PPI. For an average thesis document of 108.3 pages, it would take approximately 12.6 minutes for each document. Add approximately 5 minutes to prepare each thesis document, 2 minutes to index each thesis document and less than a minute to store each thesis document to an optical disk, it would take a total of approximately 20 minutes to prepare, scan, index and store each thesis document.

Scanning 1 thesis document every 20 minutes equals 24 thesis documents scanned in an eight hour day. Assigning one individual full time, it would take 490 days or 98 work weeks to convert the entire library of 11,750 thesis

documents. Assuming an individual hired as a GS-3 to perform the conversion, with an approximate annual salary and benefits worth $13,800, it would require a total of $26,000 to complete the initial task.

Additional conversion of 200 - 250 new thesis documents each quarter would require an individual for ten working days per quarter. Assuming the same salary requirements, it would cost an approximate $2120 per year for converting new documents.

3. Summary of Cost Analysis.

The total cost to initially implement an optical information system is $103,125, the cost of the system and disks - $77,125, plus the initial cost of converting currently stored documents, $26,000. Then to continue to convert documents as they arrive, would cost an additional $4,240 for the first two years.

VII. CONCLUSIONS, RECOMMENDATIONS

A. CONCLUSIONS

The design and implementation of a hard-copy to digital format optical information system has the potential of solving storage capacity problems not only for the NPS Knox library, but also for other document archive facilities both in the Department of Defense and other governmental and civilian agencies.

The technology to convert hard-copy documents into digital format is readily available today. The image scanning optical information system converts, stores and retrieves documents in a matter of seconds.

So the issue to determine whether or not to convert hard-copy technical documents into digital format is strictly cost. The cost of hardware and software to implement the system, the initial cost of converting currently stored documents, the cost to convert documents as they arrive, and finally, the cost of maintaining the system once it's on-line. All these costs must then be traded off for benefits in the form of space made available for other

uses, faster search and retrieval times, and an overall increase in use due to easier accessibility.

> When reviewing the performance of imaging systems in government, one can develop a cost justification based on an agency's savings in information processing costs and storing of paper. But perhaps the true bottom line should be measured in terms of service delivered to the public. (Levy, 1988, p. 6)

B. RECOMMENDATIONS

During the researcher's analysis of the thesis cage, it was noted that two documents for each thesis existed. One hard bound copy and one soft bound copy. To save space immediately, the researcher recommends removing the soft bound documents for storage elsewhere. This would free 50 percent of the space in the thesis cage. The hard bound thesis documents could then be treated as any other text in the library, being recalled if another individual needs to review a checked out thesis.

At the same time or in the future, if the decision is made to convert to an optical information system, the soft bound thesis documents could be used for scanning without interrupting the current storage system.

If the decision to save much needed floor space in the library or build now, purchasing an imaging optical information system is a highly recommended alternative. Not only to convert thesis documents, but the system could be expanded to convert other texts in the library as well.

To reduce the cost of implementing an optical information system, the researcher recommends a follow-on thesis, researching and building an in-house imaging optical information system.

A question for consideration for a follow-on thesis, would be the feasibility of scanning graphics and combining with text during thesis preparation. This would reduce the cut and paste that is currently done, reduce the overall storage requirements of a thesis, and eliminate the need to scan future theses. The final digitized copy of the thesis document could then be forwarded to the library and distributed to other government agencies at a lesser cost.

APPENDIX A

ORIGINAL PAGES USED FOR SCANNING RESEARCH

Appendix A contains the original pages from the sample thesis document used for scanning research.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

OPTICAL LASER TECHNOLOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE STORAGE AND RETRIEVAL OF INFORMATION

by

David J. Lind

June 1987

Thesis Advisor: Barry Frew

Approved for public release; distribution is unlimited

UNCLASSIFIED
URITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution is Unlimited
DECLASSIFICATION/DOWNGRADING SCHEDULE	
ERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
NAME OF PERFORMING ORGANIZATION .val Postgraduate School	6b OFFICE SYMBOL (If applicable) Code 54
7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School	
ADDRESS (City, State, and ZIP Code) nterey, California 93943-5000	7b. ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
NAME OF FUNDING/SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)
9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

TITLE (Include Security Classification)
TICAL LASER TECHNOLOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE ORAGE AND RETRIEVAL OF INFORMATION (u)

PERSONAL AUTHOR(S)
nd, David J.

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
ster's Thesis	FROM ______ TO ______	1987 June	

UPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	CD-ROM; CD ROM; CDROM; Optical Laser Disc/Disk; Information Storage; Information Retrieval

BSTRACT (Continue on reverse if necessary and identify by block number)

One of the significant problems of this "information age" is the produc-on of vast amounts of information in a form that is neither convenient r cost effective. This information is most often produced and distrib-ed on paper and the resultant effort in production, distribution and trieval is herculean. A possible solution to this, is the new optical ser technology and its use in the storage and retrieval of large amounts information. Through the use of this technology in the non-classified eas of the Department of Defense the effort in all three areas can be eatly reduced and the end user can become more efficient. In many areas DOD, the greatest benefit would be the regained space and weight asso-ated with the distribution of the manuals and other typically paper oducts on a Compact Disc - Read Only Memory (CD-ROM). One CD-ROM weighs

STRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
NAME OF RESPONSIBLE INDIVIDUAL arry Frew	22b TELEPHONE (Include Area Code) (408) 646-2924 | 22c OFFICE SYMBOL Code 54Fw

ORM 1473, 84 MAR   83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

1

Approved for public release: distribution is unlimited.

Optical Laser Technology, Specifically CD-ROM, and Its Application to the Storage and Retrieval of Information

by

David J. Lind
Lieutenant Commander, United States Navy
B.S., United States Naval Academy, 1972

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the
NAVAL POSTGRADUATE SCHOOL
June 1987

Author: David J. Lind
David J. Lind

Approved by: Barry Frew
Barry Frew, Thesis Advisor

Tung Chou
Tung Chou, Second Reader

Willis R. Greer, Jr.
Willis R. Greer, Jr., Chairman
Department of Administrative Sciences

Kneale T. Marshall
Dean of Information and Policy Sciences

3

ABSTRACT

One of the significant problems of this "information age" is the production of vast amounts of information in a form that is neither convenient nor cost effective. This information is most often produced and distributed on paper and the resultant effort in production, distribution and retrieval is herculean. A possible solution to this, is the new optical laser technology and its use in the storage and retrieval of large amounts of information. Through the use of this technology in the non-classified areas of the Department of Defense the effort in all three areas can be greatly reduced and the end user can become more efficient. In many areas of DOD, the greatest benefit would be the regained space and weight associated with the distribution of the manuals and other typically paper products on a Compact Disc - Read Only Memory (CD-ROM). One CD-ROM weighs less than an ounce and is capable of storing over 270,000 pages of text. The saved shipping and handling costs alone would be astronomically reduced not to mention the end user who would have a more effective and efficient product. The CD-ROM is designed to work as a peripheral device to a microcomputer and can therefore be made available to any user with an IBM compatible microcomputer. The

LIST OF TABLES

I. INTRODUCTION/BACKGROUND

A. GENERAL

The information age is upon us. It was reported that in 1985 the number of pages of printouts exceeded 2,000 for every man, woman, and child in America. [Ref. 1] What will we do, especially in the military, to meet this new era with the resources at hand? We cannot afford to be left behind, whether by technology or techniques. Some contemporaries have described it as an information explosion, and yet, an explosion is a singular, albeit powerful, event. The ground swell of this event is better described as a snowball rolled from the peak of the highest mountain. As it tumbles downward, it continues to increase it's momentum as it picks up more snow and velocity along it's path. From our vantage point, the slope is infinite, and although minor obstacles may be met along the way, it will continue on and on.

B. OBJECTIVE

The objective of this thesis is actually three-fold. First, the current technological capabilities in the area of optical laser research, as they apply to

not without cost and a sponsor was sought. The purchase of a CD-ROM disc drive, associated hardware and software and the cost of the services to index and master the discs, were the major costs. Naval Supply Systems Command in Washington, D.C. identified a need, a prototype application and provided support funding for this project. The hardware was purchased and the complicated process of data formatting and transfer to disc was accomplished. The actual object of the resultant demonstration was a portion of an extremely large database consisting of over 3 gigabytes (gbytes) of information composed of over 12 million total records. The prototype application dealt with approximately 360 Mbytes and slightly more than 2 million records. Although a single CD-ROM can hold up to 540 MB, the total quantity of actual data held on a disc is often much less due to the indexing requirements. Sophisticated indexing schemes can even require more space than the data itself. An example of this is Grolier Electronic Encyclopedia which requires 60 MB to accommodate the actual text of the encyclopedia and 50 MB to accommodate the sophisticated index. (See Figure 1)

The desired result of the research was to free up the two large Transaction Ledger On Disc (TLOD) disc packs each containing approximately 540 Mbytes of data

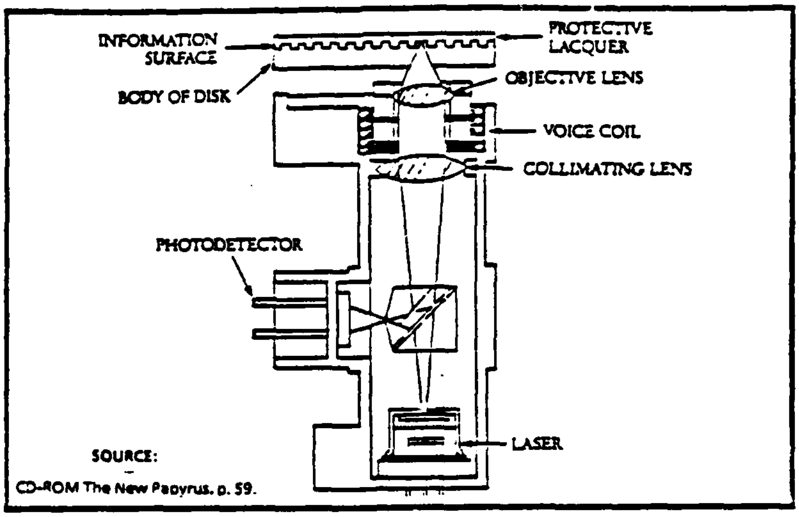

Figure 2. Optical Head Of A CD-ROM Drive

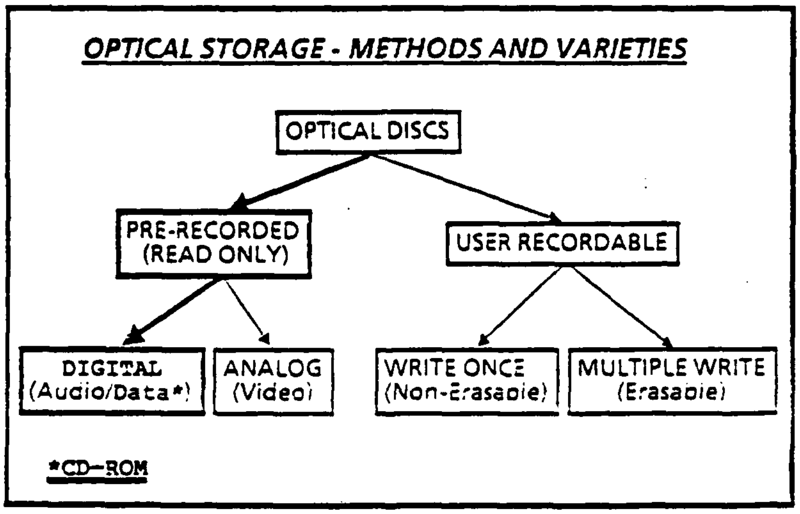

Figure 3. Optical Storage--Methods And Varieties

adjacent nonreflective pits is called a land and can also vary in its representation of data from 2 to many bits. In CD-ROM coding, a binary one is represented by the transition from pit to land and land to pit, and 2 or more zeros are represented by the distance between transitions. (See Figure 4) The resultant series lands and grooves are ultimately interpreted as one's and zero's and thus a wide variety of digitally encoded information can be stored on disc. When "reading" an optical disc, a low-power laser, senses, the presence or absence of the lands and grooves by means of reflected light energy. The small laser beam used to read back data is reflected from the lands, and scattered by the pits.

Of the prerecorded discs, the CD-ROM is the most common and draws heavily on it's predecessor, the CD-Audio Disc, for format, wide acceptance and manufacturing facilities. The recording format is a spiral groove approximately 3 miles long with a capacity of 540 MB. The tracking is maintained via the constant linear velocity (CLV) technique which requires variation of the disc rotation speed based on the distance of the read head form the center of the disc. The prerecorded disc is 4.72 inches in diameter and it's uses are primarily in the area of database distribution and permanent archival of vast amounts of

record.. The other type of prerecorded optical disc is the Optical Read Only Memory (OROM) which is slightly larger than the CD-ROM. The OROM discs are generally 5.24 inches in diameter and may be formatted with either concentric or spiral tracks. Although the capacity of the discs is very similar, OROM is often operated in a constant angular velocity (CAV) mode, thus allowing for faster access times. The typical 5 1/4" floppy disc used the CAV technique. Also, OROM may be two sided. The predominance of the CD-ROM is most probably due to its similarity to the large CD-Audio market, and the fact that CD-ROM is the only form of optical-recording that, as of this writing, has an established standard. The OROM is not expected to make a significant impact in the near future and indeed may be subsumed by the more dominant forms of optical-recording. OROM will therefore not be further addressed in this paper.

Of the two types of recordable discs, the WORM generally uses the CAV technique and the erasable disc technology is curren
tly experimenting with both techniques without a clear winner yet identified. Some of the varying physical characteristics can be seen in the following table.

APPENDIX B

RESULTS USING AN OPTICAL CHARACTER READER

This appendix includes the pages contained in Appendix A, scanned through a Kurzweil 5000 Intelligent Charater Recognition scanner. These pages are unedited to show character recognition and formatting errors. Page breaks were entered to help clarify the text that was scanned.

OPTICAL LASER TECHNOLOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE STORAGE
- AND RETRIEVAL OF INFORMATION

by

David 3. Lind

Urune 19~7

Thesis Advisor:Barry Frew

Approved for public release; distribution is unlimited

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT OOCUMENTAT1ON PAGE

Ia REP0RT SECURITY CLASSIFICATION10. RESTRICTIVE MARKINGS

Unclassif °ed

2a SECURITY CLASSIFICATION AUTHORITY3 DISTRIBUTION; AVAILABILITY OF REPORT

26.DECLASSIFICATION I DOWNGRADING SCHEDULEApproved for pu..lic release;

Distribution is Unlim°t

4 PERFORMING ORGANIZATION REPORT NUMBER(S)5 MONITORING ORGANIZATION REPORT NUMBER(S)

5a. NAME OF PERFORMING ORGANIZATION 60 OFFICE SYMBOL 7a NAME OF MONITORING ORGANIZATION

Naval Postgraduate School (If apphcable)
Code 54Naval Postgraduate School

6c. ADDRESS (City, State, anIII ZIP CooIe)lb. ADDRESS (Clry State, a- ZIP CocIe)

Monterey, California 93943-5000Monterey, California 93943-5000

8a NAME OF FUNDING/SPONSORINGB0.OFFICE SYMBOL9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
ORGANIZATION(If applicable)

Bc. ADDRESS (cry, State, a- 2'JPC'o0e)10 SOURCE OF FUNDING NUMBERS

PROGRAMPRojECTTASKWORK UNIT
ELEMENTNONONOACCESSION NO

1 1 TITLE (Include Securi Clawfica Ct on

OPTICAL LASER TEC'HNObOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE STORAGE AND RETRIEVAL OF INFORMATION (u)

12 PERSONAL AUTHOR(S)

Lind-, David 3.

'3a TYPE OF REPORT130 TIME COVERED14 DATE OF REPORT (Year, Month, Day) IS PAGE COUNT

Master's ThesisFROM ______ TO1987 June

6 SUPPLEMENTARY NOTATION

I 7C05ATI CODES10 SUBJECT TERMS (Continue on reverte if necevary and identily by block number)

F:ELDGROUPSUBGROUPCD-ROM; CD ROM; CDROM; Optical Laser Disc/Disk;
Information Storage;. Information Retrieval

'9 'ABSTRACT (Continue on reverte if necetiry and identily by block number)

One of the significant problems of this "information age" is the production of vast amounts of information in a form that is neither convenient nor cost effective. This information is most often produced and distributed on paper and the resultant effort in production, distribution and retrieval is herculean. A possible solution to this, is the ne'~ optical laser technology and its use in the storage and retrieval of' large amounts of information. Through the use of this technology in the non-classified areas of the Department of Defense the effort in all three areas can be

greatly reduced and the end user can become more efficient. In many areas '~~f DOD, the greatest benefit would be the re.gained space and weight associated with the distribution of the manuals and othe.r typically paper products on a Compact Disc - Read Only Memory (CD-ROM). One CD-ROM weighs

20 D-S'.RI3UT1ON /AVAILABILITY OF ABSTRACT 21 ABSTRACT SECURITY CLASSIFICATION

UNCLA5SIFIEDryNLIMITEDO SAME AS RpTO DT1C USERS Unc las 5 fied

22a NAME OF RESPONSIBLE INDIVIDUAL220 TELEPHONEInclude22c OFFICE SYMBOL

Barry Frew(408)eaCode Code 54Fw

00 FORM 1473, 84 MAR03 APR edition may be used until exhaustedSECURITY CLASSIFiCATIoN OF 'NIS PAGE

All other editions are ob1soIete UNCLASSIFIED

Approved for public release; distribution Is unlimited.

Optical Laser Technology, Specifically CD-ROM, and Its Application to the Storage and Retrieval of Information

by

David J. LInd
Lieutenant Commander, Vnited States Navy
B.S., Dnited States Naval Academy, 1972

Submitted in PartIal fulfillment of the-
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the
NAVAL POSTGRADDATE SCSOOL
June 1987

Author:

Approved by:

Ba y Frew, Thesis Advisor

ng Chou, Second Reader

Willis ~eer, Jr., halrman Department of Administrative
Sciences

-Kneale T. Marshall Dean of Information and Policy Sciences

3

7

LIST OF TABLES

9

I. INTRODUCTION/BACKGROUND

A.GENERAL

The information age is upon us. It was reported that in 1985 the number of pages of printouts exceeded 2,000 for every man, woman, and child in America. [Ref. 1] What will we do, especially in the military, to- meet this new era with the resources at hand? We cannot afford to be left behind, whether- by technology or techniques. Some contemporaries have- described it as an information explosion, and yet, an explosion is a singular, albeit powerful, event. The ground swell of this event is better described a- a snowball rolled from the peak of the highest mountain. As it tumbles downward, it continues to increase it's momentum as it picks up more snow and velocity along it's path. From our vantage point, the slope is infinite, and although minor obstacles may---be met along the way, it will continue on and on.

B.OBJECTIVE

The objective of this thesis is actually threefold. First, the current technological capabilities in the area of optical laser research, as they apply to

not without cost and a sponsor was sought. The purchase of a CD-ROM disc drive, associated hardware and software and the cost of the services to index and master the discs, were the major costs. Naval Supply Systems Command in Washington, D.C. identified a need, a prototype application and provided support funding for this project. 'The hardware was purchased and the complicated process of data formatting and transfer to disc was accomplished. The actual object of the resultant demonstration was a portion of an extremely large database consisting of over 3 gigabytes (gbytes) of information composed of over 12 million total records. The prototype application dealt with-approximately 360 Mbytes and slightly more than 2 million records. Although a single CD-ROM can hold up to 540 MB, tee total quantity of actual data held on a disc is often much less due to the indexing requirements. Sophisticated indexing schemes can even require more space than the data itself. An example of this is Grolier Electronic Encyclopedia which requires 60 MB to accommodate the actual text of the encyclopedia and SO MB to accommodate the sophisticated index. (See Figure 1)

The desired result of the research was to free up the two large Transaction Ledger On Disc (TLOD) disc packs each containing approximately 540 Mbytes of data

adjacent nonreflective pits is called a land and can also vary in its representation of data from 2 to many bits. In CD-ROM coding, a binary one is represented by the transition from pit to land and land to pit, and 2 or more zeros are represented by the distance between transitions. (See Figure 4) The resultant series lands and grooves are ultimately interpreted as one's and zero's and thus a wide variety of digitally encoded information can be stored on disc. When "reading" an optical disc, a low-power laser,- senses, the presence or absence of the lands and grooves by means of reflected light energy. The small laser beam used to read back data is reflected from the lands, and scattered by the pits.

Of the prerecorded discs, the CD-ROM is the most common and draws heavily on it's predecessor, the CDAudio Disc, for format, wide acceptance and manufacturing facilities. The recording format is a spiral groove approximately 3 miles long with a capacity of 540 MB.. The tracking is maintained via the constant linear velocity (CLV) technique which requires variation of the disc rotation-. speed. based on- the distance of the read head form the center of the disc. The prerecorded disc is 4.72 inches in diameter and it's uses are primarily in the area of database distribution and permanent archival of vast amounts of

records. The other type of prerecorded optical disc isthe Optical Read Only Memory (OROM) which is slightly larger than the CD-.ROM. The OROM discs are generally 5.24 inches in diameter and may be formatted with either concentric or spiral tracks. Although the capacity of the discs is very similar, OROM is often operated in a constant angular velocity (CAV) mode, thus allowing for faster access times. The typical 5 1/4" floppy disc used the CAV technique. Also, OROM may be two sided. The. predominance of the CD-ROM is most probably due to its- similarity to the large CDAudio market,. and the fact that CD-ROM is the only form of optical-recording that, as of this writing, has an - established standard. The OROM is not expected to make a significant
impact in the near future and indeed may be subsumed by the more dominant forms of opticalrecording. OROM will therefore not be further addressed in this paper.
Of the two types of recordable discs, the WORM generally uses the CAV technique and the erasable disc technology is curren tly experimenting'- with both techniques without a-clear winner yet identified. Some of the varying physical characteristics can be seen in the following table.

APPENDIX C

RESULTS USING AN IMAGE SCANNER

Appendix C contains the cover page, pages 1, 4, and 17 contained in Appendix A. The first four pages were scanned at 200 and the next four pages were scanned at 300 PPI.

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

OPTICAL LASER TECHNOLOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE STORAGE AND RETRIEVAL OF INFORMATION

by

David J. Lind

June 1987

Thesis Advisor: Barry Frew

Approved for public release; distribution is unlimited

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution is Unlimited
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School — 6b OFFICE SYMBOL (If applicable) Code 54	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b. ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION — 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
OPTICAL LASER TECHNOLOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE STORAGE AND RETRIEVAL OF INFORMATION (u)

12 PERSONAL AUTHOR(S)
Lind, David J.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM ____ TO ____	1987 June	

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	CD-ROM; CD ROM; CDROM; Optical Laser Disc/Disk; Information Storage; Information Retrieval

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

One of the significant problems of this "information age" is the produc- :ion of vast amounts of information in a form that is neither convenient ior cost effective. This information is most often produced and distrib- ited on paper and the resultant effort in production, distribution and etrieval is herculean. A possible solution to this, is the new optical aser technology and its use in the storage and retrieval of large amounts f information. Through the use of this technology in the non-classified reas of the Department of Defense the effort in all three areas can be reatly reduced and the end user can become more efficient. In many areas f DOD, the greatest benefit would be the regained space and weight asso- iated with the distribution of the manuals and other typically paper roducts on a Compact Disc - Read Only Memory (CD-ROM). One CD-ROM weighs

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassfied
22a NAME OF RESPONSIBLE INDIVIDUAL Barry Frew	22b TELEPHONE (Include Area Code) (408) 646-2924 — 22c OFFICE SYMBOL Code 54Fw

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ABSTRACT

One of the significant problems of this "information age" is the production of vast amounts of information in a form that is neither convenient nor cost effective. This information is most often produced and distributed on paper and the resultant effort in production, distribution and retrieval is herculean. A possible solution to this, is the new optical laser technology and its use in the storage and retrieval of large amounts of information. Through the use of this technology in the non-classified areas of the Department of Defense the effort in all three areas can be greatly reduced and the end user can become more efficient. In many areas of DOD, the greatest benefit would be the regained space and weight associated with the distribution of the manuals and other typically paper products on a Compact Disc - Read Only Memory (CD-ROM). One CD-ROM weighs less than an ounce and is capable of storing over 270,000 pages of text. The saved shipping and handling costs alone would be astronomically reduced not to mention the end user who would have a more effective and efficient product. The CD-ROM is designed to work as a peripheral device to a microcomputer and can therefore be made available to any user with an IBM compatible microcomputer. The

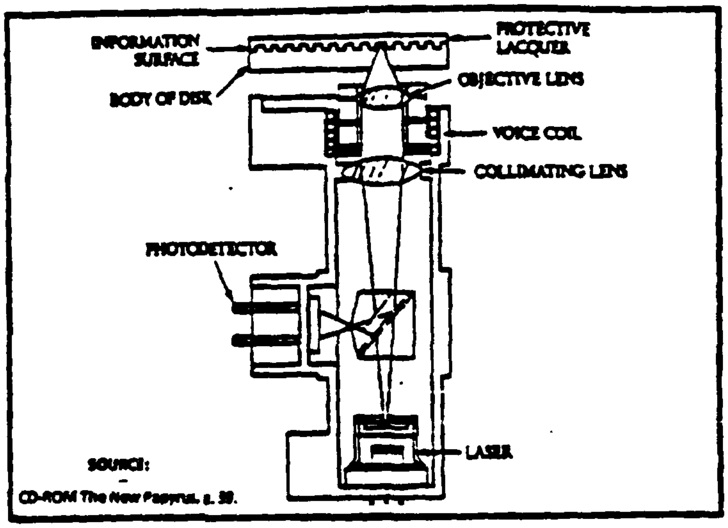

Figure 2. Optical Head Of A CD-ROM Drive

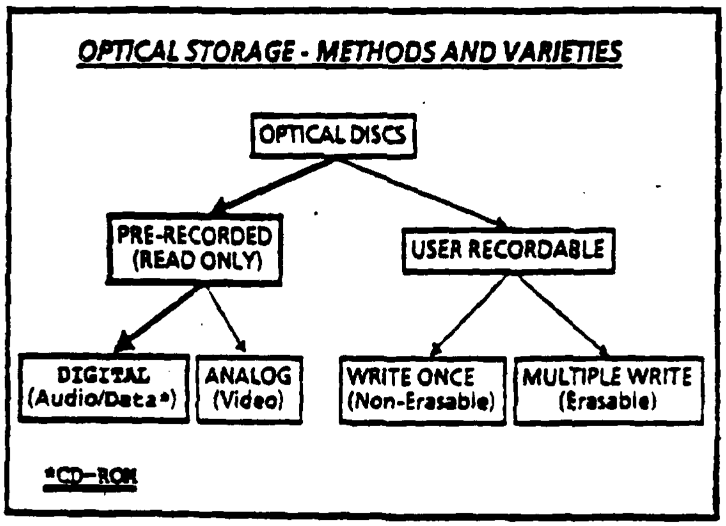

Figure 3. Optical Storage--Methods And Varieties

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

OPTICAL LASER TECHNOLOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE STORAGE AND RETRIEVAL OF INFORMATION

by

David J. Lind

June 1987

Thesis Advisor: Barry Frew

Approved for public release; distribution is unlimited

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; Distribution is Unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School; 6b OFFICE SYMBOL (If applicable): Code 54	7a NAME OF MONITORING ORGANIZATION: Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code): Monterey, California 93943-5000	7b ADDRESS (City, State, and ZIP Code): Monterey, California 93943-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO; PROJECT NO; TASK NO; WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
OPTICAL LASER TECHNOLOGY, SPECIFICALLY CD-ROM, AND ITS APPLICATION TO THE STORAGE AND RETRIEVAL OF INFORMATION (u)

12 PERSONAL AUTHOR(S)
Lind, David J.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM ____ TO ____	1987 June	

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	CD-ROM; CD ROM; CDROM; Optical Laser Disc/Disk; Information Storage; Information Retrieval

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

One of the significant problems of this "information age" is the production of vast amounts of information in a form that is neither convenient nor cost effective. This information is most often produced and distributed on paper and the resultant effort in production, distribution and retrieval is herculean. A possible solution to this, is the new optical laser technology and its use in the storage and retrieval of large amounts of information. Through the use of this technology in the non-classified areas of the Department of Defense the effort in all three areas can be greatly reduced and the end user can become more efficient. In many areas of DOD, the greatest benefit would be the regained space and weight associated with the distribution of the manuals and other typically paper products on a Compact Disc - Read Only Memory (CD-ROM). One CD-ROM weighs

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION: Unclassfied
22a NAME OF RESPONSIBLE INDIVIDUAL: Barry Frew	22b TELEPHONE (Include Area Code): (408) 646-2924; 22c OFFICE SYMBOL: Code 54Fw

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

ABSTRACT

One of the significant problems of this "information age" is the production of vast amounts of information in a form that is neither convenient nor cost effective. This information is most often produced and distributed on paper and the resultant effort in production, distribution and retrieval is herculean. A possible solution to this, is the new optical laser technology and its use in the storage and retrieval of large amounts of information. Through the use of this technology in the non-classified areas of the Department of Defense the effort in all three areas can be greatly reduced and the end user can become more efficient. In many areas of DOD, the greatest benefit would be the regained space and weight associated with the distribution of the manuals and other typically paper products on a Compact Disc - Read Only Memory (CD-ROM). One CD-ROM weighs less than an ounce and is capable of storing over 270,000 pages of text. The saved shipping and handling costs alone would be astronomically reduced not to mention the end user who would have a more effective and efficient product. The CD-ROM is designed to work as a peripheral device to a microcomputer and can therefore be made available to any user with an IBM compatible microcomputer. The

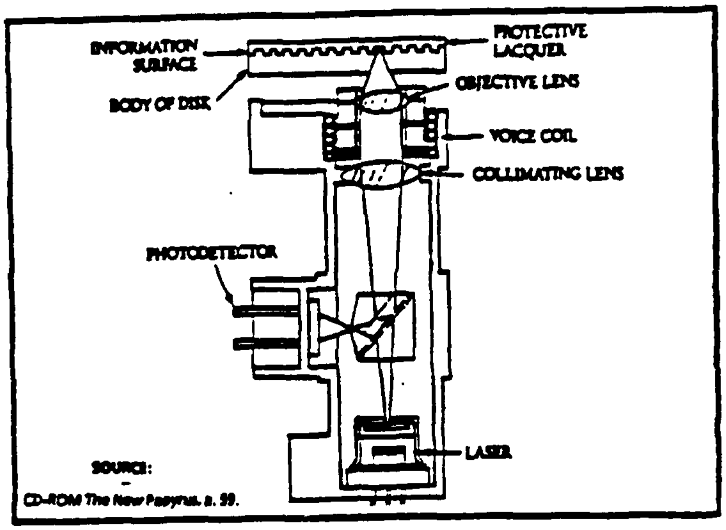

Figure 2. Optical Head Of A CD-ROM Drive

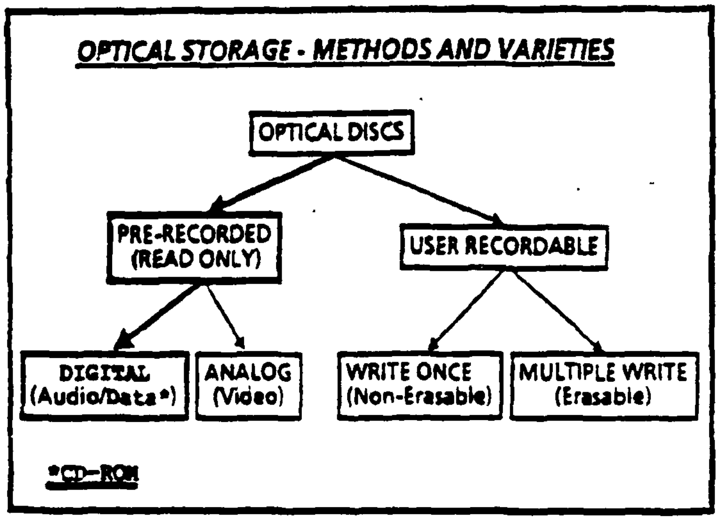

Figure 3. Optical Storage--Methods And Varieties

GLOSSARY

Analog--Analog data is a representation of information by a signal that varies in proportion to the amount of the original information. Thus, the size of a signal, such as light, is expressed by another signal, an electrical voltage, that is proportional to the amount of light reflected.

ANSI--American National Standards Institute

Application Development--Customer software developed according to the user's specification that can include user interface, data presentation and integration of the information product into existing applications.

ASCII--American Standard Code for Information Interchange. It is the standard table of 7-bit digital representations used to transmit information to a printer, other computers, or other peripheral devices.

Binary--Binary data is a representation of numerical information that uses only two expressions. These are, numerically, the digits "1" and "0" or, electronically, "on" and "off." Thus, the on/off representation allows electronic storage and manipulation of the information.

Bit--Binary digit. The smallest part of information in binary notation. A bit is written as either 1 or 0 and represents either the on or off variation of voltage.

Board--A printed-circuit board, or card, that mounts onto the physical chassis of a computer or peripheral and holds the chips and associated. wiring. Other cards may be plugged into this board.

BPI--Bits per inch is usually used to describe the electronic representation on a video screen; a bit is frequently equivalent to a pixel.

Buffer--An auxiliary storage area for data. Many peripherals have buffers used to temporarily store data that will be used as time permits.

Byte--A group of eight bits of digital data which is processed together. A byte can have 256 (or 28) possible combinations of 8 binary digits.

CAV--Constant Angular Velocity. A technique that spins a disc at a constant speed, resulting in the inner disc tracks passing the read/write head more slowly that the outer tracks. this results in numerous tracks forming concentric circles with the storage density being the greatest on the inner track. (See also CLV).

CCD--Charged Coupled Device is a device composed of a row of several thousand small photocells. Each pixel on the output image corresponds to a photocell. A CCD is actually a one-chip microcircuit.

CCITT--Acronym for the French name of the Consultive Committee on International Telephone and Telegraph. CCITT issues the standards for data compression techniques such as CCITT Group 3.

CD--Compact Disc - See CD-ROM

CDI--Compact Disc Interactive. Physically identical to the CD-ROM disc, however, with emphasis on the interactive presentation of video, audio, text and data. A self-contained multimedia system expected to operate in conjunction with home entertainment equipment.

CD ROM--See CD-ROM

CDROM--See CD-ROM

CD-ROM--Compact Disc - Read Only Memory. A computer peripheral capable of storing large amounts of data which are placed on the disc at the time of manufacture.

Checksum--A method of checking the accuracy of a character transmitted, manipulated, or stored. The checksum is the

result of the summation of all the digits involved. Used for error detection vice error correction.

Chip--The term applied to an integrated circuit that contains many electronic circuits. A chip is sometimes called an IC or an IC chip. The name is occasionally applied to the entire integrated circuit package.

CIRC--Cross-Interleaved Reed-Solomon Code. The only error correction scheme used with CD Audio, and the first layer used with CD-ROM. It is implemented in the hardware, and uses two independent R-S codes to achieve an error rate of 1 uncorrected error per 109 bytes.

CLV--Constant Linear Velocity (as opposed to CAV). Used with CD-ROM to keep the data moving past the optical head at a constant rate. In order to accomplish this, the rotational speed of the disc must vary, decreasing as the head moves from the inner tracks toward the outer perimeter. The range is approximately 500 to 200 rpm for a CD-ROM disc drive.

Code--A method of representing data in a form the computer can understand and use.

Command--A code that represents an instruction for the computer.

CRC--Cyclic Redundancy Code. ECC algorithm for the checking of CD-ROM after error correction is performed--only capable of error detection.

Density--The closeness of space distribution on a storage medium such as a disc.

Digital--Digital data is a representation of information by numerals. Thus, the size of the electrical voltage is expressed as numbers: that is, in digits.

Disc Preparation--Providing certified tapes and shipping containers for customer data. Scanning input tapes for data integrity and cleaning up minor problems, building a directory (High Sierra or customer), putting the data in proper format for the mastering center user.

DPI--Dots per inch refer to the dots, or spots, of ink placed on paper by a printer; each may be composed of more than one pixel.

DOS--See Disk Operating System.

Double-Density--This term is most often applied to the storage characteristics of disks, and generally refers to the density of the storage of bits on the disk surface on each track.

DRAW--Direct Read After Write. A write once optical disc technology (See also WORM), an error control technique; however, it is unable to be used with CD-ROM.

EBCDIC--Extended Binary Coded Decimal Interchange Code. An 8-bit code developed by IBM, and used primarily by IBM and its compatibles. The code is used to represent 256 numbers, letters and characters in a computer system. (See also ASCII)

ECC--Error Correction Coding. The application or addition of data to the original data in order to provide a means of correction when an error in the original data is detected.

EDAC--Error Detection and Correction. Redundant information which is calculated according to certain algorithms used to detect and correct errors when data is read.

EDC--Error Detection Code. The application of redundant data to the original data in order to detect errors.

GB--See Gigabyte.

Gbyte--See Gigabyte.

Giga--1,000,000,000.

Gigabyte--1,000 megabytes, or 1 billion (109) bytes.

Glass Master--The original glass disc upon which the digital information is burned with a laser. From it are formed the "stampers" which in turn are used to produce the numerous discs, usually by an injection molding process.

Hardware--The physical computer and all of its component parts, as well as any peripherals and interconnecting cables.

HeadCrash--When the read-head contacts the magnetic surface of the disk--a highly undesirable occurrence.

High Sierra Group--An ad hoc working group of CD-ROM service companies, vendors, and manufacturers which has been a prime source of activity in the setting of standards for CD-ROM data format and compatibility. The group was named after its first meeting place--the High Sierra Hotel at Lake Tahoe. The group first met in 1985.

IC--Integrated Circuit.

Indexing--The actual processing of all records according to the layout and the building of the index file. Indexes permit the computer to rapidly locate data without searching through the full body of data. Generally, a data item is searchable only if it is indexed.

Indexing Set Up--Tape handling, resource allocation and - loading the layout programs on the indexing system.

Instruction--A program step that tells the computer what to do for a single operation in a program.

Interface--A device that serves as a common boundary between two other devices, such as two computer systems or a computer and peripheral.

Jewel Box--The plastic container in which the CD-ROM disc is generally stored.

Jukebox--See Optical Jukebox.

K--Abbreviation for Kilo.

KB--See Kilobyte.

Kbyte--See Kilobyte.

Kilo--A prefix meaning (1) 1000 when used in a mathematical expression; or (2) 1,024 210 when used as a unit measure in computers. As an example, 16K would equal 16 times 1,024 or 16,384.

Kilobyte--A unit of measure in computers that equals 1024 bytes.

LAN--Local Area Network.

Land--The reflective area between two adjacent non=reflective pits on a disc. The transition from pit to land or land to pit represents a binary 1. (See also Run).

M--Abbreviation for Mega.

Magneto-Optic--A form of erasable media that stores information in the form of vertically oriented magnetic domains.

Mastering--The entire process involving the scheduling of the mastering center, managing artwork and packaging issues and Q.A.ing all replicas for data integrity and readability.

MB--See Megabyte.

Mbyte--See Megabyte.

Mega--1,000,000.

Megabyte--1,000 Kilobytes, or 1 million (106) bytes.

Metal Mother--The negative mold created from the glass master which is in turn used to stamp the numerous discs. Often called a "stamper".

Micron--One square micron, the area occupied by 1 bit on a CD-ROM. One millionth of a meter.

Microsecond--One 1/1,000,000th of a second.

Millisecond--One 1/1,000th of a second.

MO--See Magneto-Optic.

MS-DOS--The disk operating system used with IBM computers and their compatibles.

OCR--Optical Character Recognition. Generally used in reference to a device capable of scanning printed material into a digital form.

ODS--Optical Digital Data Disc

Optical Jukebox--A store and read mechanism capable of storing and accessing multiple CD-ROMs. Accessing is generally accomplished by mechanical means afterwhich the discs are placed on a single reader (disc drive) for use.

OROM--Optical Read Only Memory

Photoconverter--See CCD.

Photocell--A photocell is an electronic component which changes a light signal into an electrical signal by photoelectric conversion. A photocell is only a few microns square.

Pit--The microscopic depression in the reflective surface of a disc. The pattern of pits represents the data being stored on the disc. (See also ""land''). The light from the laser used to read the data is reflected back from the lands, but scattered by the pits. A typical pit as about the size of a bacterium - 0.5 by 2.0 microns.

Pixel--A pixel (picture element) is the smallest controllable element of an image. As resolution (the number of pixels per inch) increases, pixel size decreases and

details are more accurately represented. Pixels are usually square, but they may be rectanqular or round. The shape is determined by the optical system of the device.

PPI--Pixels per inch.

Platter--Generally used in reference to the larger (12'') optical discs. Sometimes in reference to a single layer in a magnetic disc pack.

RAM--Random Access Memory. Semiconductor memory circuits used to store data and programs in information processing systems.

Resolution--Resolution is defined as the number of pixels read or displayed per inch (PPI), both horizontally and vertically.

R/W/E--/Read/Write/Erase--An alternative title for erasable discs.

Run--The distance between transitions either from land to pit or pit to land. The distance represents two or more zeros (See also Land).

SCSI--Small Computer Systems Interface--A complete 8bit parallel interface bus structure with rates up to 4 Mbytes/sec. that is subordinate to the rest of-the system architecture. Up to 8 systems and peripherals may be connected to the same bus.

Software--A general term that applies to any program (set of instructions) that can be loaded into a computer from any source.

SPI--Spots per inch. See DPI.

Stamper--See Metal Mother

Substrate--The base material form which a disc is made, generally a strong and transparent polycarbonate plastic.

Tbyte--Terabyte or 1,000 gigabytes. (1012)

Track--A linear, spiral or circular path on which information is placed, or found. The portion of a disk that one read/write head passes over to extract data. Track density is measured in tpi (tracks per inch).

WORM--Write Once Read Many (occasionally seen as Write Once Read "Mostly" or "Multiple").

LIST OF REFERENCES

Alter, Allan, "The Unpapering of America," *CIO Magazine*, v. 2, No. 1, pp. 16-26, October 1988.

"CD-ROMs: The Laser's Edge in Data Storage," *Mechanical Engineering*, pp. 50-55, April 1987.

Chavez-Simmons, Martha, "Understanding OCR Technology," *Micro User's Guide*, pp. 18-23, Summer 1988.

Dukeman, John, "Optical Disk - A Technology on the Move," *Modern Office Technology*, pp. 82-88, June 1988.

Grigsby, Mason, "Optical Disk - Vision to Payoff," *Modern Office Technology*, pp. 60-68, November 1988.

Kapoor, Ajit, "Electronic Imaging and Information Processing," *The Office*, p. 28, December 1988.

Lambert, Steve, and Ropiequet, Suzanne, "CD ROM the New Papyrus," *Microsoft Press*, p. 68, 1986.

Lesher, Donald, "Making the World Safer Place for Trees," *CIO Magazine*, v. 2, No. 1, p. 33, October 1988.

Levine, Ron, "Optical Storage Comes of Age," *Dec Professional*, pp. 49-59, November 1988.

Levy, I., "Controlling Paperwork with Image Processing," *Government Technology*, v. 1, No. 6, pp. 5-6, November/December 1988.

Lind, David J., *Optical Laser Technology, Specifically CD-ROM, and its Application to the Storage and Retrieval of Information*, Master's Thesis, Naval Postgraduate School, Monterey, California, June 1987.

Matlin, Mark, "Image Compression for Document Storage," *ESD: The Electronic System Design Magazine*, pp. 75-76, July 1988.

McNaul, Jim, "Scanners Bring Grayscale Processing to Desktop Publishing," *Micro User's Guide*, pp. 27-30, Fall 1988.

Rosch, Winn, "Worms for Mass Storage," *PC Magazine*, pp. 135-166, 23 June 1987.

Imaging - Primer Series, Wang Laboratories, Inc., 1987.

Xerox 7650 Pro Imager Reference Manual, Xerox, September 1987.

BIBLIOGRAPHY

Ambrosio, Johanna, "IRS Experiment Puts Tax Returns On Optical Disk," *Government Computer News*, pp. 39-40, 22 July 1988.

Alter, Allan, "Image Meets Reality," *CIO Magazine*, v. 2, No. 1, pp. 28-33, October 1988.

Anderson, Julie, "Succeeding at CD-ROM," *PC Tech Journal*, pp. 9-10, October 1988.

Black, Lauren, "Image Scanners," *Info World*, pp. 41-52, 11 July 1988.

Bridge, Raymond "Casting Databases into Plastic," *CD-ROM Review*, November/December 1987.

Brindza, Stephen, "Managing Text at a Glance," *Modern Office Technology*, pp 102-104, May 1988.

Brunet, James, "Scanners Cling To PC Software," *Unix World*, pp. 79-86, October 1988.

Chickering, John E., "The Advent of the Paperless Ship," *Naval Engineers Journal*, pp. 226-236, May 1988.

"Colorado Starts Optic Conversion," *Government Technology*, v. 1, No. 6, p. 14, November/December 1988.

Danca, Richard, "AGA Inc. Shows Rewritable Optical Disks for PCs," *Government Computer News*, p. 13, 22 July 1988.

Devoy, Jim C., "Media, File Management Schemes Facilitate WORM Utilization," *Computer Technology Review*, p. 48-49, Fall 1988.

Dinman, S., "Image/Text Scanners," *Government Computer News*, v. 7, pp.71-77, 22 July 1988.

Forbes, Jim, "AI Software Lets Optical Scanners Read Entire Page," *PC Week*, p. 6, 15 August 1988.

Francis, B., "PC Back-up's Optical Understudy," *Datamation*, pp. 57-60, 15 December 1988.

Fusco, Joseph, "Image Processors Need Faster Data Transfer," *Computer Technology Review*, pp. 20-21, June 1988.

Goff, Leslie, "Kodak Slated To Add Bus to Optical Jukebox," *Management Information Systems Week*, p. 14, 19 September 1988.

Greenis, Mark, "Redefining the Scope of Office Automation," *Government Technology*, v. 1, No. 6, pp 12-19, November/December 1988.

Haber, Lynn, "Optical-Disk Storage Market Just Keeps on Inching Along," *Digital News*, pp. 61-63, 19 September 1988.

Hecht, Jeff, "Lasers Store A Wealth of Data," *High Technology*, pp. 408-415, May/June 1982.

Helgerson, Linda, "In Search of CD-ROM Data," *PC Tech Journal*, pp. 67-75, October 1988.

Helgerson, Linda, "CD-ROM Publishing Strategies," *PC Tech Journal*, pp. 53-63, October 1988.

Helliwell, John, "Optical Overview: What's Coming in CD-ROMs and WORMs," *P C Magazine*, pp. 149-164, 14 October 1986.

Henricks, Mark, "WORM Squirms for a Place in the Sun," *PC World*, pp. 74-76, September 1988.

Hosinski, Joan, "Agencies Urged to Ease Into Optical-Disk Applications," *Government Computer News*, p. 31, 19 December 1988.

Hosinski, Joan, "Agencies Watch Progress of PTO's WORM System," *Government Computer News*, pp. 10-14, 9 January 1989.

Kalstrom, D., "Optical Disks Find Increased Applications," *Computer Technology Review*, p. 30, November 1988.

Kaplan, A., "Scanner Well Suited for Local Goverment," *Government Technology*, v. 1, No. 6, p. 33, November/December 1988.

Kellis, Mark, "CD-ROM Revolutionizes Information Distribution in the Federal Government," *Micro User's Guide*, pp. 14-16, Fall 1988.

Lacy, John, "The Benefits of Image-Based Document Management," *The Office*, pp. 132-137, January 1989.

Langworthy, George, "Insights to Optical Storage," *Storage*, pp. 13-14, May 1988.

Levine, M., "Going Beyond Text to Images Is Next Step," *Digital Review*, pp. 59-61, 9 January 1989.

"Mass Storage Options," *Computer Buyer's Guide and Handbook*, v. 4, pp. 14-15, June/July 1986.

McCormick, John, "Suppliers Benchmark Claims Aren't Always Valid," *Government Computer News*, v. 7, No. 15, pp. 74-77, 22 July 1988.

Miles, J.B., "Remembrance of Things to Come," *Computer Decisions*, pp. 105-107, 14 January 1986.

Miyatake, Masaya, "Image Scanner Sales Promise to Double in 1989," *Computer Technology Review*, pp. 22-25, November 1988.

Moberg, Dick, "Videodiscs and Optical Data Storage," *Byte Magazine*, pp 400-405, June 1982.

Mortensen, Erik, "The CD-ROM Debate: What is Its Potential Value," *The Office*, pp. 71-72, May 1988.

"Optical Data Storage: The State of the Art," *Computer Buyer's Guide and Handbook*, v. 4, pp. 12-13, June/July 1986.

Page, Bruce, "Lan Optical Disks," *LAN Magazine*, pp. 63-68, August 1988.

Rangel, Diane, "Putting Texas on Disc," *CD-ROM Review*, pp. 40-42, November/December 1987.

Ross, David, "Omnifont Character Recognition Provides Flexible Data Handling," *Computer Technology Review*, Winter 1987.

Ruster, Allen, "Capacity Differences Complicate Optical Disk Choices," *Government Computer News*, p. 44, 22 July 1988.

Schlosser, Phillip, "Electronic Paper Does Away With Keyboard," *Computer Technology Review*, p. 24, November 1988.

Schnaidt, Patricia, "Worm Networks - Maimonides Hospital Reduces Paper With Optical Disks," *LAN Magazine*, pp.. 70-74, August 1988.

Sehr, Barbara, "High Noon for CD-ROM," *Datamation*, pp. 79-88, 1 November 1986.

Simmons, A., "The Paperless Office - Dream or Vision," *Government Technology*, v. 1, No. 6, pp. 1-6, November/December 1988.

Simpson, David, "Erasable Optical Disks: When, What...Why?," *Mini-Micro Systems*, pp. 42-60, December 1987.

Smith, Steven, "The Final Steps: Look Before You Reap," *CD-ROM Review*, November/December 1987.

Stanton, Tom, "Page-To-Disk Technology," *PC Magazine*, pp. 128-177, 30 September 1986.

Stanton, Tom, "Scanners Take Off," *PC Magazine*, pp. 185-196, 13 October 1987.

Stork, Carl, "CD-ROM: The New Distribution Medium," *Micro User's Guide*, pp. 25-28, Summer 1988.

Strukhoff, Roger, "Iris Eyes - Intermedia Prize," *CD-ROM Review*, pp. 18-27, May 1988.

Sullivan, Judith, "IRS Pilots Optical Image Storage Retrieval Systems," *Government Computer News*, v. 5, No. 18, 10 October 1986.

Teja, Edward, "Object-orientation and Images: Keeping Up With Improving Output Devices," *Hardcopy*, p. 31, June 1988.

Teja, Edward, "Move Down the Wire," *Hardcopy*, pp. 32-42, June 1988.

Watson, Paula, "CDROM Catalogs-Evaluating LEPAC and Looking Ahead," *Online*, pp. 74-79, September 1987.

Willett, Shawn, "Omnipage Makes Low Cost Desktop Scanning A Reality," *Computer Technology Review*, p. 24, November 1988.

Wilson, A., "The Big Easy - Cramming Over 2000 Images on a Hard Disk," *ESD: The Electronic System Design Magazine*, p. 24, September 1988.

Wilson, A., "High-Resolution Scanners Come of Age," *ESD: The Electronic System Design Magazine*, pp. 17-18, October 1988.

Worsley, Charles, "Reading and Writing and Discmatic," *CD-ROM Review*, pp 44-46, November/December 1987.

DISTRIBUTION LIST

		No. Copies
1.	Library, Code 01422 Naval Postgraduate School Monterey, CA 93943-5002	2
2.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
3.	Chief of Naval Operations Director Information Systems (OP-945) Navy Department Washington, DC 20350-2000	1
4.	Director, Naval Data Automation Command Washington Navy Yard Washington, DC 20374-1662	1
5.	Commanding Officer Navy Fleet Material Support Office P.O. Box 2010 (Attn: Code 9513) Mechanicsburg, PA 17055	1
6.	Barry A. Frew Administrative Sciences Department Code 54FW Naval Postgraduate School Monterey, CA 93943-5000	6
7.	Carl R. Jones Administrative Sciences Department Code 74 Naval Postgraduate School Monterey, CA 93943-5000	1
8.	LCDR David J. Lind USN Naval Data Automation Command (Code 30) Washington Navy Yard Washington, DC 20374-1662	1

9. LT Mike Pafford USN 1
Commander
Naval Security Group Support Activity (GX)
3801 Nebraska Ave. N.W.
Washington, DC 20393-5220

10. MAJ Paul W. LeBlanc USMC 1
Marine Corps Central Design & Programming Activity
Marine Corp Combat Development Command
Quantico, VA 22134

11. LCDR Robert R. Taylor USN 2
Patrol Squadron Seventeen
FPO San Francisco 96601-5910

12. MAJ Elbert T. Shaw USA 1
1278 Spruance Rd.
Monterey, CA 93940

13. Office of the Secretary of Defense 1
DEERS Support Office
Attn: CAPT Dellaporta
2511 Garden Road, Suite 260
Monterey, CA 93940-5331

14. Command Officer 1
Fleet Material Support Office
Attn: LCDR P.R. Richey
5450 Carlisle Pike
Mechanicsburg, PA 17055